

D 233 TTGGAATGGCCGAGNCGTTCGTAATTCCTGCTTCCACGACATCATTAAC 292
OY 516 TTGAGATGGCCGAGGTTCTCTGTAAGTTCTCTTCTGCTGCTGCTAAGCAGTAT
D 293 CAAGTGGATGATGGAAGCTTCTGTCGTAAGATGATTCATTCCTCAAAATATGTCG 352
OY 576 AGCGTCTTGAAGTGGCAAGCTT--TCCG-GCCA-GAATTCCTGAC-TGTC-T--TGA-G 626
D 353 CAATAGATATATCAGGGAATATCAGATTCGACATTTGATTAATTCGGAGCTTTGT 412
OY 627 CACTGCTGCTCCAGCTCAGCAGCTGCTCTTCTGACTTTGAGCAGCTGGAAGATCCTG 686
D 413 GATATGATCCAGATGGAATATATATTTGATGATGGAATGATGATGAGAGTA 472
OY 687 GAAGCTATTTCCAGGGAATATATATGCTGATGCTGGAATGCTACTCTGAACAC 746
D 473 TTGCTGACTTTCATACGACGCTGACGACCAATTTCCACACATTAATCTTTGGCGG 532
OY 747 TTGTTGAATTTGTAAGAAATGTAAGGAGAGCGCTTCCCGACGACATCATGCGAGG 806
D 533 AATGCTGTCACAGGTGAATGTTGAGAGCTCATACTATCCGCGCGGAGCTGTGAAG 592
OY 807 AATGTGTAACAGAGAGATGTTGAGAGAGCTCATCTTCTGGGCTGACATCATCAA 866
D 593 GTTGGCATCGCTCCGGGGTGGGTGTCACACAGCGCTGAAGAGCTGTTGGATACCA 652
OY 867 GTGGGAATTTGGCGAGGCTGTGTGTAACCTGGAAGAAACTGGAAGGGGTTATCCA 926
D 653 CAGCTGATGCGGCTGACTGATGATGTCAGACGACATCTGATGCTCAAGGGGACGATG 712
OY 927 CAGCTGAGCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 986
D 713 AGCAGCGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 772
OY 987 TCAGATGAGAGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1046
D 773 TTGCTTATGATGAGAGATTTATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 832
OY 1047 TTGCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1106
D 833 AAGATGCGAAGAGTATTAATTTATGATGATGATGATGATGATGATGATGATGATG 892
OY 1107 AGGATGCGAAGAGTATTAATTTATGATGATGATGATGATGATGATGATGATGATG 1166
D 893 TATCAGGATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 952
OY 1167 TATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1226
D 953 AGAGTACGATATGCGACAGATACAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1012
OY 1227 AAGAGATGATGGAACATATCCATCCGAGACATCTTGAAGAGGATCCGCTCTAC 1286
D 1013 TATCAGGACCACTATGCTCAAAAGACTCTCAAAAGGCGCAGCTTGTCCGCTGACG 1072
OY 1287 TATGTTGGACAGCTAAGCTCAAAAGAGTGAAGAGGAACTATCTTATCCGAGTACC 1346
D 1073 CAACAG 1078
OY 1347 CAGCAG 1352

RESULT 3
LOCUS ECAE00119 12434 bp DNA BCT 02-SEP-1997
DEFINITION Escherichia coli K-12 MG1655 section 9 of 400 of the complete
genome
ACCESSION AF000119 U00096
NID 91786283
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
Eubacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
1 (bases 1 to 12434)

AUTHORS
Blattner, F.R., Plunkett III, G., Bloch, C.A., Perna, N.T., Burland, V.,
Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
Mau, B. and Shao, Y.
The complete genome sequence of Escherichia coli K-12
Science 277 (5331), 1453-1474 (1997)
97426617
MEDLINE
2 (bases 1 to 12434)
Blattner, F.R.
AUTHORS
Blattner, F.R.
TITLE
Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
3 (bases 1 to 12434)
Blattner, F.R.
AUTHORS
Blattner, F.R.
TITLE
Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
COMMENT
The E. coli K-12 sequence and its annotations have been updated.
All of the ambiguous residues in our original submission have been
resolved, and mis-assemblies in two repetitive regions have been
realigned. The annotations have been improved and updated as well.
With this release we begin designating a version number for the
annotated sequence, to assist in keeping track of corrections,
updates, and other changes. This is version M52 (SRPT. 02. 1997).
In addition, a revised notation has been instituted which assigns
each gene (protein- or RNA-encoding) a unique numeric identifier
beginning with a lowercase 'b' (in the 'label' field); this will
remain constant through further updates, gene identifications, etc.
This sequence was determined by the E. coli Genome Project at the
University of Wisconsin-Madison (Frederick R. Blattner, director).
Supported by NIH grants HG00301 and HG01428 (from the Human Genome
Project and NCHGR). The entire sequence was independently
determined from E. coli K-12 strain MG1655.
Predicted open reading frames were determined using Genemark
software, kindly supplied by Mark Borodovsky, Georgia Institute of
Technology, Atlanta, GA, 30332.
e-mail: mark@embarc.gatech.edu
Open reading frames that have been correlated with genetic loci are
being annotated with CG Site Nos., unique ID nos. for the genes in
the E. coli Genetic Stock Center (CGSC) database at Yale
University. Kindly supplied by Mary Berlyn. A public version of the
database is accessible (<http://cgsc.biology.yale.edu>).
Annotation of the genome is an ongoing task whose goal is to make
the genome sequence more useful by correlating it with other data.
Comments to the authors are appreciated. Updated information will
be available at the E. coli Genome Project's World Wide Web site
(<http://www.genetics.wisc.edu>).
FEATURES
source
1..12434
/organism="Escherichia coli"
/sub_strain="K-12"
/db_xref="taxon:562"
61..1212
/gene="ftsZ"
61..1212
/gene="ftsZ"
/note="c383; 99 pct identical to ftsZ_ECOLI SW: P06138"
/codon_start=1
/label="b0095"
/product="cell division protein FtsZ"
/db_xref="PID:91786284"
/transl_table=11
/translation="MEPEWELNDAYIKIVIGVGGGAGNAEVMERIEGVFEAVNT
DAQALRKAVGOTIOIGSITGKIGAGANPEVRGAADRDALAALEGAMVRIA
GNGGCTGCAAPVAVAKDGLIVAVYTRFNPENKRRKRAFOGTEELKRVDS
TITPDKRLKIVGRISLIDARGANDVAKGVQVIGIAELITPGALMADFADVRVLS
ENGVAMGSGVGEDRAEAGAEAMASPLLEDIDLSARGALVITGTFDLRDEFE
IVGNITRAFASDNAFTVIGTSLDPMNDELRTVAVATIGMDKREPEITLVTKVOQOP

gene
1313. .2230
/gene="lpxc"
1313. .2230
/note="0305; 100 pct identical to lpxc_ECOLI SW: P07652;
alternate gene name envA"
/codon_start=1
/label-b0096
/product="UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine
deacetylase"
/db_xref="PID:g1786285"
/transl_table=11
/translation="MKQRLTKRIVQAGVGLHGKQKVTLLIRPAPANTGVIRRD
NPPYDPADAKSVSDTLCCTCYVNEHVRISTVHLLAALAGIDIVLEVNAPEP
IMGSAPEVYLLDAGIDELNCCKKVRKREYRVSDGMAEFKPYNSDTEID
FNHPAIDSSNQRYAMNSADAFMRQISRATPEYDLYSGKGLCGSPCAIYV
DDYRVHAIKTRVIRHLSFAPAPQTLPAVESLPLQAQHLALDLTLALTLQEGTSPSK
GYRIDYHETPPQAKFTPWVISOAGIRAGPQRLT"
3035. .5740
/gene="seca"
3035. .5740
/note="0901; 99 pct identical to SECA_ECOLI SW: P10408"
/codon_start=1
/label-b0098
/product="preprotein translocase SecA subunit"
/db_xref="PID:g1786287"
/transl_table=11
/translation="MLIKLLKYFGSSNDRTLRMRKVNIINAMEPEMEKLSDEELK
GKAEFRARLEKEGVLENLIPFAVAVREASRKYGRHFDVQLGGMVNERCIAM
RTGSKTLTATLPAYNALTGKGVHTVNDYLAORDAENNRPLFEFLGTVGINTLG
MPAPKREAVAAADITVTGNNYGFYLRDMNAPSPEERYQRKHYALVDVSDTLIDE
ARPLILISGPAEDSSEMYKRVNKTIIPHLINQEKDSTFPGEHGFSDERSQVNLTE
RGVLIELLVKESIMDEGSLSYSPANIMLNHTALRNHALTRVDYIVKGEVY
IVDEHTRTQGRMSDGLHQAVAKGVQIQNDQVLASTFQNYRELKLGAMG
TADTEAFESISYKLDIVVTPNPMRKDLPDYVTEAEKQIATEDIKERAKQ
PVLVGTISIKSELVSNELTKAGIKHNVLNKKFANFANAIVAOGYPAATVATNMG
RGDIVIGGSQMAVALENPLAEOIKIKADQVRRDAVLEAGGLITITHEHESPR
IDNULRSGRQDAGSSRYLSHEDLMATFADRSYGMRKIKGMPGRIETHPWT
KALNAQKVESRNFDRKQULEYDVANQRRATYQNNELLDVSQVSETINSIRD
VFATIDAYIPQSLMEMDIPGLQERLKNDFDLDLIAWLKEPLHETLERIL
AQSIEYQREKEVYVGAEMRHFKEKVMQLQTLDSKMLAMDRLGILRGAGAOQ
PKQIEKRESFMFAAMELSLKEYEIVSLISQVQVPEVELEBOGRMEARLQAQMO
LSHODDSAAAALAAQTGERKVRNDPCGSGSKKTKQCHGRQ"
5800. .6189
/gene="mutr"
5800. .6189
/note="mutr"
/EC_number="3.6.1.-"
/function="involved in the GO system responsible for
removing oxidatively damaged guanine from DNA and the
nucleotide pool"
/note="0129; 100 pct identical to MUTR_ECOLI SW: P08337"
/codon_start=1
/label-b0099
/product="7,8-dihydro-8-oxoguanine-triphosphatase"

gene
/db_xref="PID:g1786288"
/transl_table=11
/translation="MKKIQAVGCIITNENNEIFITRRADAHANKLEPPGKIEGE
TPEQAVVRELDQEVGITPQHFSLFEKLETFEPDRHITLFWLWVERWEGEPWKEGPG
EWSVLVGLNADDFPPANPEVIAKRL"
complement(6320. .6544)
/note="f44; 61 pct identical (1 gap) to 21 residues of an
approx. 216 aa protein KTHY_HUMAN SW: P23919"
/codon_start=1
/label-b0100
/db_xref="PID:g1786289"
/transl_table=11
/translation="MAIFPKAMGAKNQSSDICLMPHVGILRRGQRIRHLYQMSDA
"
complement(6405. .6602)
/gene="yacg"
complement(6405. .6602)
/gene="yacg"
/note="f65; 100 pct identical to YACG_ECOLI SW: P36681 (50
aa) but contains 15 additional N-terminal residues"
/codon_start=1
/label-b0101
/db_xref="PID:g1786290"
/transl_table=11
/translation="MSETITVNCPTGKTVWGEISPRFQSKRCQLDLGEWMAE
KRIPSSGDLSESDMSEBPKO"
complement(6612. .7355)
/gene="yacf"
complement(6612. .7355)
/gene="yacf"
/note="f247; 98 pct identical to YACF_ECOLI SW: P36680 (78
aa) but has 169 additional N-term residues"
/codon_start=1
/label-b0102
/product="hypothetical protein in mutr-guac intergenic
region"
/db_xref="PID:g1786291"
/transl_table=11
/translation="KQTVLFEPHPLNEKRWIRIEFLQQLTVNPIYDHGALHFF
RNSVELDVEFEREVRTELKELDRQQRKIQTVIGVGVDSRIELIQLKAAGSVL
ISAPRIGQFREDRLIALVROSLIPGGCCSEPLPYLHILHPQORDSOVETWTAS
LNPQLQALTVLDLIROSAPFERKQTSILNGFYQDNGSDADLLRLNLSLDSQLYQISGH
KSNFAIRFNPDLDENGQVPERDLFELAC"

Query Match 11.0%; Score 203; DB 13; Length 12434;
Best Local Similarity 60.6%; Pred. No. 1.02e-153;
Matches 623; Conservative 0; Mismatches 402; Indels 3; Gaps 3;
Note: remainder of annotations omitted.

Db 8205 TATTGAAGAGATCTGAAGTATAGSTTTTAAAGACGTTCTGATCCGCCCTAAAGCCTGCAC 8264
Qy 344 TATTGACACAGATCTGAAGCTGAACTCAAGAGATGCTTTTGAGGCCCAAGCAGCTAC 403
Db 8265 TCTTAAAGCGCTTCGATGTGAACGTGAACGTCATTCACCTTCAAAATTCAGTCA 8324
Qy 404 CTTTAAGTCTCGAAGTAGGATGATCTCACAAAGTCCCTTTCATTTGGAACCTCAAGCA 463
Db 8325 GAGCTGTCGCCGCTGCCGATTAATGCCGCAAAATAGACACCGTAGACATTTCTAT 8384
Qy 464 GACATCTCTGGGTTTCCCATCATCTGCTGCCAATATGATATCTGGGACACCTTTGAGAT 523
Db 8385 GGCCTCTGCGCTGCTCTTTGATATTTTGAATGCTGTCATTAACACATATCTGCGA 8444
Qy 524 GGCACAGGTTCTCTGTAAGTCTCTCTCTCTACACGCTGCATTAAGCATAATGCGTGT 583
Db 8445 AGATGCGACGCTTTATCAACAATTTCTCCGCTGATGTGCTGAATAATGTGATGTTTC 8564
Qy 584 TCACTGCGACAGAGTTTCTGCGCAGAACTCTGACTGCTTGAGAGATC-TG-GCGGCCAGC 641
Db 8505 TACCGTACGCTGATGCGGATTTGCAAAAACATACAGATTTCTGCACTGAAACCGGCG 8564
Qy 642 T-CAGGCAAGGCTCTTCTGACTTTGAGACGCTGGAACAGATCTCTGGAAGCTATTTCCCA 700

D	b	8965	ATTAACTTGTTGTAATGTAGACTGGCCGAATGGTATTATCCCAACAACTTCGTGCAGTTGCT	8624
O	y	701	GGTGAATAATATATAGCCTGGATGTGGCAATATGGCTAATCTCGAACTTGTGTAATTTGT	760
D	b	8625	TGCAGAACGCCGCGAAGCCGGCGCCGACCAAAACCATTTGTGCTGGTAACGTAGTACTGG	8684
O	y	761	AAAAGATGTACGGAGAAGCGCTTCCCACAGCACACATCATATGCGAAGGAAATGTGTAAACAG	820
D	b	8685	TGAATATGTAGAGAGCTTATCTCTCAGGTGGCCATATCGTTAAAGTTGGCATTGGCCCC	8744
O	y	821	AGAAATGGTAGAAGAGCTCATCTCTTCGTGGGGCTGCATCATCAAAGTGGAAATTGGGCC	880
D	b	8745	AGGTTCTGTTTATACAATCGCTGCCTCAAAAACAGCCCGGTTTTCCGGCAACTTTCGCGCT	8804
O	y	881	AGGGCTGTGTGTATCTACTCTGGAAGAAATCGAATGGAGGGGTATCCACAGCTCAGCGCAGT	940
D	b	8805	AATGAAATGTGCGCATGCTGCGACGCGTCTGGCGGAATATGATCTCAGAGATGGTGCCTG	8864
O	y	941	GATGAGATGTGCAAGATGCTCTCATATGGCTCAAAAGCCACATATTTCCAGATGGAGTTG	1000
D	b	8865	CACCACGCCGGCGGATGTGGCGAAAGCCTTTGGCGGGGTGCCGATTTCTGTCATGCTTGG	8924
O	y	1001	CAGCTGTCTCGGGGATGTGGCCAAAGCCTTTTGGGGCAGAGACTGACTTCGTATCTGGG	1060
D	b	8925	CGGATCTGTGGGGGCCACGAAAGAGAGCGGCGTCCATCTGTTAGAGAAACGGCGCAAA	8984
O	y	1061	TGGATCTGTGGCTGGGACGATAGTCAGTGGTGTGAGCTCATATCGAGAGGATGGCAAGAA	1120
D	b	8985	ATTATGCTGTTCTACGCGATAGCTCCGAGTCTGCGATGAAGAAACGTACGTTGCGGGCGT	9044
O	y	1121	GTACAACTCTTCTATGATGATGAGTTCTGAANAATGGCCATGAAAGATATGCTGGGGGGCT	1180
D	b	9045	TGCGGAATATCGCGCAGCAGAGAGTAAAAACGTTAAGCTGCGCTGCGAGGCCCGGTTGA	9104
O	y	1181	GGCTGATACAGAGGCCCTCAAGGAGAAAGACAGTGAAGTCCTTTAAAGGAGATGTGA	1240
D	b	9105	AAATACGGCGAGATATTTTGGGCGGCCCTGGCTGAGTTTACATATCGTTGGGGCTTC	9164
O	y	1241	ACATACCATCCGAGACATCTCTAGAGAGGATCGCTTACGTGACCTATGTGGGAGCG	1300
D	b	9165	ACGCGCTAAGAGAGCTGCACCAAGCGCACCAAGTTATCTGTGTGCAAGAACAAAGACCG	9224
O	y	1301	TAACTCAAAAGATTAGACGAGAACTACCTCATTCGAGTACCCACAGAGTAAATCC	1360
D	b	9225	CATCTCA 9232	
O	y	1361	AATCTTCA 1368	

RESULT	4			
LOCUS	ECGUACGM	1991 bp	DNA	BCT 03-DEC-1992
DEFINITION	Escherichia coli quac gene for guanosine 5'-monophosphate oxidoreductase (Ec 1.6.6.8).			
ACCESSION	X07917 M33020			
NID	942615			
KEYWORDS	GMP reductase; quac gene; guanosine 5'-monophosphate oxidoreductase; oxido-reductase; reductase.			
SOURCE	Escherichia coli.			
ORGANISM	Eubacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.			
REFERENCE	1 (bases 1 to 1991)			
AUTHORS	Andrews, S.C.			
JOURNAL	Direct Submission Submitted (15-JUN-1988) Andrews S.C., The Dept of Microbiology, Sheffield University, Western Bank, Sheffield, S10 2TN, United Kingdom			
REFERENCE	2 (bases 1 to 1991)			
AUTHORS	Andrews, S.C. and Guest, J.R.			
TITLE	Nucleotide sequence of the gene encoding the GMP reductase of Escherichia coli K12			
JOURNAL	Biochem. J. 255 (1), 35-43 (1988)			

FEATURES	MEDLINE 89061679
source	location/Qualifiers 1..1991
	/organism="Escherichia coli" /strain="K12" /db_xref="taxon:562" /clone="pGS89" /map="2.6 min"
-35_signal	25..30 /note="promoter 2"
repeat_unit	44..49 /note="inverted repeat A"
-10_signal	50..55 /note="promoter 2"
prim_transcript	61..>1250 /note="transcript from promoter 2"
repeat_unit	62..67 /note="inverted repeat A"
-35_signal	98..103 /note="promoter 1"
repeat_unit	119..130 /note="inverted repeat B"
-10_signal	122..127 /note="promoter 1"
-35_signal	126..131 /note="promoter 3"
misc_feature	128..134 /note="discriminator sequence promoter 1"
repeat_unit	132..147 /note="inverted repeat B"
prim_transcript	135..>1250 /note="transcript from promoter 1"
-10_signal	148..153 /note="promoter 3"
misc_feature	155..161 /note="discriminator sequence promoter 3"
prim_transcript	160..>1250 /note="transcript from promoter 3"
RBS	198..201 /note="ribosome binding site"
CDS	210..1250 /codon_start=1 /product="GMP reductase (AA 1-346)" /db_xref="PID:g42616" /db_xref="SWISS-PROT:P15344" /transl_table=1 /translation="MRIEDLKGFKYVLIIRPKSTLKSRSDVLEKQFFPKHSGQSM SGVPIIANDMTVGFESMASLASFDTLTAHKHY SVEWQAFINSSADVLKRVVVS TGSIDADFEKTKQITLDINPALNFVCIDVANGYSEHFVOFAKAREAMPRTKICGNVVS TGECECELILSGADIVKVGIGSGVCTRYATGVGYPOLSAVICADAHLGMIYSV DGGCTTGGDYAKAFARADFYMLGMLGHSRGGRIYEENGEKPLPYGSSSESAMKR HNGGVAYRAAEGTKVTLPLRGPVENTARDILGSLBSACTYVGASRLKELIKRTITIR VOEDENRIENNL"
repeat_unit	285..291 /note="inverted repeat C"
repeat_unit	301..307 /note="inverted repeat C"
repeat_unit	620..627 /note="inverted repeat D"
repeat_unit	644..651 /note="inverted repeat D"
repeat_unit	908..915 /note="inverted repeat E"
repeat_unit	923..928 /note="inverted repeat F"
repeat_unit	932..939 /note="inverted repeat E"
repeat_unit	933..938 /note="inverted repeat F"
repeat_unit	979..987 /note="inverted repeat G"
repeat_unit	997..1005 /note="inverted repeat G"

Best Local Similarity	60.5%;	Pred. No. 7.10e-138;
Matches	622; Conservative	0; Mismatches 400; Indels 6; Gaps 5;

Matches	622;	Conservative	0;	Mismatches	400;	Indels	6;	Gaps	5;
---------	------	--------------	----	------------	------	--------	----	------	----

QY 1361 AATCTCA 1368

Nucleotide sequence encoding the dihydrolipoamide acetyltransferase

- | | | | |
|---|--|---|---|
| JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE | Component
Eur. J. Biochem. 133 (3), 481-489 (1983)
83234434
4 (sites)
Stephens,P.E., Lewis,H.M., Darlison,M.G. and Guest,J.R.
Nucleotide sequence of the lipamide dehydrogenase gene of
Escherichia coli K12
Eur. J. Biochem. 135 (3), 519-527 (1983)
84004369
5 (sites)
Richard,C., Richard,F., Martin,C., Haziza,C. and Patte,J.C.
Regulation of expression and nucleotide sequence of the Escherichia
coli dcpd gene
J. Biol. Chem. 259 (23), 14824-14828 (1984)
85054973
6 (sites)
Broome-Smith,J.K., Edelman,A., Yousif,S. and Spratt,B.G.
The nucleotide sequences of the ponA and ponB genes encoding
penicillin-binding protein 1A and 1B of Escherichia coli K12
Eur. J. Biochem. 147 (2), 437-446 (1985)
85127060
7 (sites)
Coulton,J.W., Mason,P., Cameron,D.R., Carmel,G., Jean,R. and
Rode,H.N.
Protein fusions of beta-galactosidase to the ferrichrome-iron
receptor of Escherichia coli K-12
J. Bacteriol. 165 (1), 181-192 (1986)
86085668
8 (sites)
Breton,R., Sanfacon,H., Papayannopoulos,I., Biemann,K. and
Lapointe,J.
GlutamyI-tRNA synthetase of Escherichia coli. Isolation and primary
structure of the gltX gene and homology with other aminoacyl-tRNA
synthetases
J. Biol. Chem. 261 (23), 10610-10617 (1986)
86278132
9 (sites)
Koster,W. and Braun,V.
Iron hydroxamate transport of Escherichia coli: nucleotide sequence
of the fnuB gene and identification of the protein
Mol. Gen. Genet. 204 (3), 435-442 (1986)
87014116
10 (sites)
Chye,M.L. and Pittard,J.
Transcription control of the arop gene in Escherichia coli K-12:
Analysis of operator mutants
J. Bacteriol. 169 (1), 386-393 (1987)
87083395
11 (sites)
Ben-Bassat,A., Bauer,K., Chang,S.Y., Myambo,K., Boosman,A. and
Chang,S.
Processing of the initiation methionine from proteins: properties
of the Escherichia coli methionine aminopeptidase and its gene
structure
J. Bacteriol. 169 (2), 751-757 (1987)
87109068
12 (sites)
Coulton,J.W., Mason,P. and Allatt,D.D.
fnuB and fnuH genes for iron (III)-ferrichrome transport into
Escherichia coli K-12
J. Bacteriol. 169 (8), 3844-3849 (1987)
87279948
13 (sites)
Tabor,C.W. and Tabor,H.
The speBped operon of Escherichia coli. Formation and processing
of a proenzyme form of S-adenosylmethionine decarboxylase
J. Biol. Chem. 262 (33), 16037-16040 (1987)
88058963
14 (sites)
Gebhard,W., Schreitmuller,T., Hochstrasser,K. and Wachter,E.
Complementary DNA and derived amino acid sequence of the precursor
of one of the three protein components of the inter-alpha-trypsin
inhibitor complex | JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE | FEBS Lett. 229 (1), 63-67 (1988)
88152237
15 (sites)
Mellano,M.A. and Cooksey,D.A.
Nucleotide sequence and organization of copper resistance genes
from Pseudomonas syringae pv. tomato
J. Bacteriol. 170 (6), 2879-2883 (1988)
88227880
16 (sites)
Andrews,S.C. and Guest,J.R.
Nucleotide sequence of the gene encoding the GMP reductase of
Escherichia coli K12
Biochem. J. 255 (1), 35-43 (1988)
89061679
17 (sites)
Sung,Y.C. and Fuchs,J.A.
Characterization of the cyn operon in Escherichia coli K12
J. Biol. Chem. 263 (29), 14769-14775 (1988)
89008347
18 (sites)
Lipinska,B., Sharma,S. and Georgopoulos,C.
Sequence analysis and regulation of the htrA gene of Escherichia
coli: a sigma 32-independent mechanism of heat-inducible
transcription
Nucleic Acids Res. 16 (21), 10053-10067 (1988)
89057448
19 (sites)
Liu,J.D. and Parkinson,J.S.
Genetics and sequence analysis of the pcnB locus, an Escherichia
coli gene involved in plasmid copy number control
J. Bacteriol. 171 (3), 1254-1261 (1989)
89155419
20 (sites)
Xie,Q.W., Tabor,C.W. and Tabor,H.
Spermidine biosynthesis in Escherichia coli: promoter and
termination regions of the speD operon
J. Bacteriol. 171 (8), 4457-4465 (1989)
89327165
21 (sites)
Lindquist,S., Galleni,M., Lindberg,F. and Normark,S.
Signalling proteins in enterobacterial Ampc beta-lactamase
regulation
Mol. Microbiol. 3 (8), 1091-1102 (1989)
90113890
22 (sites)
Roncero,M.I., Jepsen,L.P., Stroman,P. and van Heeswijk,R.
Characterization of a leuA gene and an AKS element from Mucor
circinellus
Gene 84 (2), 335-343 (1989)
90128278
23 (sites)
Honore,N. and Cole,S.T.
Nucleotide sequence of the arop gene encoding the general aromatic
amino acid transport protein of Escherichia coli K-12: homology
with yeast transport proteins
Nucleic Acids Res. 18 (3), 653 (1990)
90174991
Erratum: [Nucleic Acids Res 1990 Mar 11;18(5):1332]
24 (sites)
Surlin,B.P., Watson,J.M., Hamilton,W.D., Economou,A. and Downie,J.A.
Molecular characterization of the nodulation gene, nodT, from two
biovars of Rhizobium leguminosarum
Mol. Microbiol. 4 (2), 245-252 (1990)
90251164
25 (sites)
Kang,P.J. and Craig,E.A.
Identification and characterization of a new Escherichia coli gene
that is a dosage-dependent suppressor of a dnaK deletion mutation
J. Bacteriol. 172 (4), 2055-2064 (1990)
90202727
26 (sites)
Wurgler,S.M. and Richardson,C.C.
Structure and regulation of the gene for dGTP triphosphohydrolase |
|---|--|---|---|

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is F13A2, 800 bp overlap; 3' cosmid is F20D3. Actual start of this cosmid is at base position 197 of CELF32D1; actual end is at 39235 of CELF32D1. This cosmid lies in an unanchored cluster, the orientation of which is unknown.

NOTES:

Coding sequences below are predicted from computer analysis, using the program GeneFinder (P. Green and L. Hillier, ms in preparation).

FEATURES

source

1. .39235

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db_xref="taxon:6239"

/chromosome="V"

/clone="F32D1"

3013. .4116

/gene="F32D1.8"

join(3013. .3199,3850. .3921,3971. .4034,4107. .4116)

/gene="F32D1.8"

/codon_start=1

/evidence=not_experimental

/db_xref="PID:g2291235"

/translation="MIGYLISLQAWLKRVFVPTKTNIAKGVADSPFSVSEISERP

MIKEPDEQHDDEWEHXRARIQDLRLRVLPPIRGRKTLLEQGSADSDDEEDMSDSE

PAIKCKRLG

4568. .5597

/gene="F32D1.7"

join(4568. .5108,5187. .5459,5506. .5597)

/gene="F32D1.7"

/note="coded for by C. elegans cDNA yk137a12.5; coded for

by C. elegans cDNA yk184f6.5; coded for by C. elegans cDNA

yk140e9.5; coded for by C. elegans cDNA CEMS24F; coded

for by C. elegans cDNA yk140e9.3; coded for by C. elegans

cDNA yk189g11.3; coded for by C. elegans cDNA yk184f6.3;

coded for by C. elegans cDNA CEMS30F"

/codon_start=1

/db_xref="PID:g2291229"

/translation="MADEDDPCGGGKRNPRKRCQPDVDPEDSEPDLDGPPPPGRGS

STYDAVAROTAMEFRKARSDERPAQDQSVAVAPPTSAVTPFANLPVSEQL

EDSLGVRVRVRLPRSRYPVANDPDKFSAYSGSSQOAVPTRTSKISESGSGS

GGSGNDSSGSGPTBMSGOVPTSGPPPPPPAGHONPFLGQSTWMSSESTPT

OGVSPRTSKISKKAHQAGLHSGSESDTSGESKPSKQPPKPSKKNPNAP

GPSRRDDDDGVGGSSGSSSESL"

5977. .6809

/gene="F32D1.6"

join(5977. .6063,6298. .6396,6447. .6554,6609. .6809)

/gene="F32D1.6"

/codon_start=1

/evidence=not_experimental

/db_xref="PID:g2291230"

/translation="MSYDAFYSTAFPIPGPNLGYTAYSPGISFNDSPASSVNT

PSYFSGSIHSRNSIAMSISHSILSNDISGTSSTPSPTPEIHSNAPKKTIP

LYLSDDDECDKSRKRIRSRKSKSRGIRKSRKKAADYRKIKMLGTFEGKATVL

KLKK"

complement(7083. .8832)

/gene="F32D1.9"

complement(join(7083. .7783,7842. .8060,8112. .8403,

8454. .8735,8785. .8832))

misc_feature

gene

CDS

9374. .9375

/note="Str1 trans-splice site; see EST yk82f3.5"

9393. .11266

/gene="F32D1.5"

join(9393. .9635,9693. .10235,10976. .11266)

/gene="F32D1.5"

/note="Similar to GMP reductase; coded for by C. elegans

cDNA yk77a7.3; coded for by C. elegans cDNA CEMS93F;

coded for by C. elegans cDNA yk82f3.3; coded for by C.

elegans cDNA yk61d12.3; coded for by C. elegans cDNA

yk64f12.3; coded for by C. elegans cDNA yk74e3.3; coded

for by C. elegans cDNA yk80e1.3; coded for by C. elegans

cDNA yk90g10.5; coded for by C. elegans cDNA yk77a7.5;

coded for by C. elegans cDNA yk82f3.5; coded for by C.

elegans cDNA yk81a3.5; coded for by C. elegans cDNA

yk80e1.5; coded for by C. elegans cDNA yk74e3.5; coded for

by C. elegans cDNA yk64f12.5; coded for by C. elegans cDNA

yk61d12.5"

/codon_start=1

/db_xref="PID:g2291231"

/translation="MPRIENEPKUDFKDYLLRPKSTLSKRVADVELDEHYVRNSKAT

YTGVPVNASMDVTGTFEMAAALNNKRIFTTHKHYSVDEMKAFASAPDFENLAI

SSGISNDMTKNTVTTELTPOLKYLICLDPAVDESEFVEKRVREARPKRTIMAGNV

VTGEMVEELISADIVKYGIGSGVCTRRKAGVGPQLSAVLECADAHGLNGVAM

SDGSGNPDVAKAFGADFWMIGLFGHGOOSGDLEHNNKRRKFLFGMSDQAM

KRHGSAVEYRSEKGTVTIPYRGDVNGVYDILGSRACCTTGAKHKLAKRATF

IRYTGQTDMYVEFVTPAPRSK"

16666. .18013

/gene="F32D1.4"

join(16666. .16765,16874. .16996,17195. .17290,17337. .17422,

17473. .17622,17837. .18013)

/gene="F32D1.4"

/codon_start=1

/evidence=not_experimental

/db_xref="PID:g2291238"

/translation="MOKIPYNNISIPFKNONCTKITEKTSIAOQLSPRPPTII

QOQPYSVAPRIPIQOQHNSQOQIPYFNNSNAPPSSTSSYNAAPSQNAAPQ

ASKKEISADFTDLIOKSKPEITDKMDNDVDVNAISNPEISLOAKSMKRNIRRAQ

AAATEKAGCSESLQOIMEVSRHAMSNSVSKLISRGAKEGYNFEDYCSQD

FSYLISNIFCRVLEGOICLAYEN"

18289. .18522

/note="C. elegans telomere-like hexamer repeat"

19190. .30869

/gene="F32D1.3"

join(19190. .19329,22428. .22604,24581. .24666,25151. .25389,

27083. .27612,28007. .28258,28980. .29553,30543. .30869)

/gene="F32D1.3"

/note="contains similarity to TPR domains; coded for by C.

elegans cDNA yk28d9.3; coded for by C. elegans cDNA

yk28d9.5"

/codon_start=1

/db_xref="PID:g2291234"

/translation="MKKERKROKSVPTLISHIPYOYLVALEFASIVRTLNADPVY

DDRPALITNDVIGRTPWRSILYHNDWNGPDIQSSHSYRFLITASPRIOAGVL

KALFPHGVNICMINSMLYKAROMRINKGNFSLISALIRPCHPTSEAYCSYGR

ADLLSTMLILALITCHTSDSTRTVLSTLATAETGILILPLITLDTVEKTSNT

KQRRVYATLILALCYLRLSINNEQSKRNPNLAHEPNSLRLALTDVLEKTSNT

HMLDIYFKLSTYDMSMDALPKVESLDSRIILFTVIGVEHLKLLGLNNSNS

HNERSLFLPALTPTTHILLSNLTGVFAARILYLVAVACILAAFLAECSSRS

SPKSVLYITLILSTFIRTMORVDMMKEESFKSALVENPDKAMLOSTYTAALH

YELAKYHYRQALRCOGNLADAWYNILILSKTSNNNGSHCTOMALOSTYTAALH

NLALLHDGHEHMAAFSHDKCLNNTGEFKFTHVNRKTOATCAFKGRILQSGNFH

VALENELAKLGLGPHFEHTPSVNLNGMCYNLGDQSEKFEFGAILENHNSLYLT

MAHLKIRONSFEVENLLRKAMTLAPESVYLVONIALAEFHMONYNSILFTYRKALHL

Note: remainder of annotations omitted.

```

Query Match      10.0%; Score 185; DB 15; Length 39235;
Best Local Similarity 71.8%; Pred. No. 6,03e-137;
Matches 305; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

Db 9811 TTCGCACTCAAGTACATCTGTCTTGAAGTGGCCAGGAGATATCTGATCTTCGTGG 9870
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 694 TTCGCCAGCTGAATATATATGCTGTGATGTGGCAATGGCTACTGTAACACTTTGTG 753
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9871 AGTTATTCGCCGCTGTGTCAGAGAGCTACCCGAGACACATCATGCGCGGAAAGCTTG 9930
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 754 AATTGTAAAGATGTAGGGAAGCGCTTCCGCCACACATCATGCGAGGAGATGTGG 813
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9931 TCACCGAGAGATGTGTAGAGAGCTGATTTCTGTGCGCGGACATTTCAAGTGGAA 9990
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 814 TAACAGGAGAGATGTGTAGAGAGCTCATCTTTCTGGGCGTGCATCATCAAGAGGGAA 873
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9991 TCGGACGGGATCGGTTTGACACCACTCGCAGAGAGCCGGAGTTGGATACCCAGAGCTTA 10050
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 874 TTGGGCCAGGCTCTGTGTGTACTACTCGGAAGAAAGACTGGAGTGGGATCCACAGCTCA 933
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10051 GTGCCGCTCTGAGTGCCTGTGATGCCGCTCATGTCTCATGAGATGATGATG 10110
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 934 GCGCAGTGTATGATGTGTGATGTGCTGCTCATGCGCTCAAGGCCACATCATTTAGATG 993
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10111 GTGATGCAAGCAATCTTGAGATGTGGCAAGGCTTTTGGAGCTGGAGCAGATTTGCTCA 10170
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 994 GAGGTTCGACGCTGCTGTGGGATGTGGCCAAAGGCTTTTGGGAGAGAGCTGCTGCTGA 1053
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10171 TGATTTGAGAGACTTCTTCCTGACACAGATCAGATGAGAGAGATCTCATGAGCATATG 10230
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1054 TGTGTGGTGGCTGCTGCTGTGGGACAGTGTGATGCTCAATCGAGAGGATG 1113
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10231 GAAG 10235
      |||||
Qy 1114 GCAAG 1118
      |||||

RESULT 7
LOCUS S73035S9 601 bp DNA PRI 10-JUN-1992
DEFINITION guanosine monophosphate reductase [human, Genomic, 601 nt, segment
9 of 9].
ACCESSION S73075
NID 9241139
KEYWORDS 9 of 9
SEGMENT human.
SOURCE Homo sapiens
ORGANISM Eukaryotes; Mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 601)
AUTHORS Kondoh,T., Kanno,H., Chang,L. and Yoshida,A.
TITLE Genomic structure and expression of human guanosine monophosphate
reductase
JOURNAL Hum. Genet. 88 (2), 219-224 (1991)
MEDLINE 92098099
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI g1bseq 73075] from the original journal article.
This sequence comes from figure 5.

FEATURES
SOURCE
1. .601
Location/Qualifiers
/db_xref="taxon:9606"
/organism="Homo sapiens"
order(S73035:1..279,S73043:1..159,S73045:1..130,
S73047:1..232,S73049:1..158,S73053:1..170,S73060:1..96,
S73066:1..207,1..601)
/gene="guanosine monophosphate reductase"

BASE COUNT 131 a 167 c 140 g 163 t

Query Match 5.1%; Score 94; DB 22; Length 601;

```

```

Best Local Similarity 76.1%; Pred. No. 2.13e-54;
Matches 137; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Db 36 CAGAGCCTCTGAGGTAAGACTGTGGAAGTTCCTTACAAAGAGATGTGAAACACTAT 95
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1190 CAGAGCCTCAGAGGAAGAGAGTGAAGTTCCTTTAAAGAGATGTGGAATACAT 1249
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 96 CCTGATATTTCTGGGGAGCTGAGGTCCAGTGCAGTACCTAGTGGGGCCCAACTCAA 155
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1250 CCGAGACATCTTACAGGAGATCCGCTCTAGTGTACTATGTGGACACTTAAGCTCAA 1309
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 156 GGAGCTCAGAGAGGCAACATCATCCGGTGTGACCCAGCAGCAACACCGTGTAC 215
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1310 AGATTGAGCAGGAGAACTTCTTATCCAGTACCCAGCAGGTGAATCAATCTTAC 1369
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 8
LOCUS HS467D16 143583 bp DNA PRI 18-JAN-1998
DEFINITION Human DNA sequence from PAC 467D16 on chromosome 6p22.3-24.1.
Contains the 3' part of the SCAL (ataxin-1) gene with a
poly-glutamine (CAG repeat) polymorphism, the 3' part of the GMPR
(GMP reductase, Guanosine 5'-monophosphate oxidoreductase) gene,
ESTs and an STS with a polymorphic CA repeat.
AL009031
6p22.3-24.1; ataxin-1, CA repeat polymorphism; CAG repeat
polymorphism; GMPR; Guanosine 5'-monophosphate oxidoreductase;
SCAL.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotes; Metazoa; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 143583)
Tubby,B.
REFERENCE Direct Submission
AUTHORS Submitted (19-JAN-1998) sanger.ac.uk/HGP/Chrf/ Sanger Centre,
Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
IMPORTANT: This sequence is the entire insert of clone 467D16.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variations annotated may not be found in the sequence submission
corresponding to the overlapping clone as we submit sequences with
only a small overlap as described above.
This sequence was generated from part of bacterial clone contigs of
human chromosome 6, constructed by the Sanger Centre chromosome 6
mapping group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chrf/
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key
The true left end of clone 467D16 is at 1 in this sequence. The
true right end of clone 467D16 is at 143583.
467D16 is from the library RPEC13 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong.
For further details see http://bacpac.med.buffalo.edu/.

FEATURES
SOURCE
1. .143583
Location/Qualifiers
/db_xref="taxon:9606"
/organism="Homo sapiens"
/chromosome="6"
/map="6p22.3-24.1"
/clone="467D16"
/clone_1id="RPEC13"
99. .252
repeat_region
/note="MER20 repeat: matches 164. .4 of consensus"
prim_transcript complement(507..931)
/note="match: ESTs T05971"
repeat_region 1860..1929

```

```

repeat_region 21019..21317 /note="AluB repeat: matches 3..302 of consensus"
repeat_region 22600..22841 /note="MTR1A2 repeat: matches 374..128 of consensus"
repeat_region 26002..26353 /note="HEIC repeat: matches 371..1 of consensus"
repeat_region 26942..27243 /note="AluB repeat: matches 302..1 of consensus"
repeat_region 27997..28318 /note="AluB repeat: matches 1..300 of consensus"
repeat_region 29551..29808 /note="MIR repeat: matches 62..217 of consensus"
prim_transcript 30254..30403 /note="match: ESTs T19981 T19980"
repeat_region 30465..30582 /note="2 copies of 59 mer 86 % conserved"
prim_transcript 31248..31452 /note="match: EST AA364443"
repeat_region 31686..31946 /note="AluX repeat: matches 392..44 of consensus; incomplete repeat"
repeat_region 32104..32159 /note="3 copies of 22 mer 94 % conserved"
repeat_region 32473..32512 /note="20 copies of 2 mer 95 % conserved"
repeat_region 33075..33376 /note="AluX repeat: matches 302..1 of consensus"
repeat_region 34397..34699 /note="AluX repeat: matches 1..301 of consensus"
repeat_region 34734..34804 /note="L1P45 repeat: matches 820..890 of consensus"
repeat_region 35273..35564 /note="AluY repeat: matches 1..294 of consensus"
repeat_region 36546..36664 /note="MTR2 repeat: matches 146..24 of consensus"
repeat_region 39656..39703 /note="MTR5 repeat: matches 53..5 of consensus"
repeat_region 39803..40042 /note="L1P1 repeat: matches 902..656 of consensus"
repeat_region 40000..40058 /note="L1P47 repeat: matches 698..640 of consensus"
repeat_region 41306..41743 /note="L1P82 repeat: matches 658..204 of consensus"
repeat_region 42306..42607 /note="AluX repeat: matches 1..302 of consensus"
repeat_region 42615..42664 /note="5 copies of 10 mer 90 % conserved"
repeat_region 44012..44314 /note="AluX repeat: matches 1..301 of consensus"
repeat_region 45528..45619 /note="AluY repeat: matches 3..301 of consensus"
repeat_region 45721..45874 /note="MTR5B repeat: matches 178..3 of consensus"
repeat_region 45886..45964 /note="MTR5A repeat: matches 188..109 of consensus"
repeat_region 45887..45985 /note="MTR5A repeat: matches 17..109 of consensus"
repeat_region 46182..46482 /note="AluS9 repeat: matches 300..1 of consensus"
repeat_region 46671..47200 /note="MTR4A repeat: matches 536..1 of consensus"
repeat_region 48173..48469 /note="AluB repeat: matches 1..297 of consensus"
repeat_region 48772..49074 /note="AluB repeat: matches 297..1 of consensus"
repeat_region 49573..49680 /note="MIR repeat: matches 183..75 of consensus"
repeat_region 50354..50652 /note="AluSC repeat: matches 299..1 of consensus"
repeat_region 50772..50887 /note="MIR repeat: matches 227..101 of consensus"

```



```

repeat_region complement(21348..21649)
/rpt_family="AluSg"
repeat_region 21781..21812
/rpt_family="AT_rich"
repeat_region 22678..22872
/rpt_family="MIR"
repeat_region complement(23592..23697)
/rpt_family="LINE2"
repeat_region 23733..23866
/rpt_family="AluJb"
repeat_region complement(25793..26168)
/rpt_family="MSRA"
repeat_region complement(27863..27978)
/rpt_family="MER30"
repeat_region complement(28081..28129)
/rpt_family="AT_rich"
repeat_region complement(28223..28436)
/rpt_family="AluJo"
repeat_region complement(28499..28642)
/rpt_family="AluJo"
repeat_region complement(29156..29371)
/rpt_family="MLT1C"
repeat_region complement(29409..29694)
/rpt_family="AluSg"
repeat_region complement(29695..29878)
/rpt_family="MLT1C"
repeat_region 30397..30442
/rpt_family="LINE2"
repeat_region complement(30765..30814)
/rpt_family="GAAAn"
repeat_region 31035..31123
/rpt_family="LINE2"
repeat_region complement(31291..31811)
/rpt_family="L1PB3"
repeat_region 31846..32144
/rpt_family="AluJb"
repeat_region 32227..32384
/rpt_family="MIR"
repeat_region complement(33427..33668)
/rpt_family="MLT1A2"
repeat_region complement(36830..37191)
/rpt_family="THE1C"
repeat_region complement(37717..37748)
/rpt_family="(CAAA)n"
repeat_region complement(37750..38052)
/rpt_family="AluJb"
repeat_region complement(38309..38411)
/rpt_family="(GAAAn)"
repeat_region 38806..39127
/rpt_family="AluJo"
repeat_region 39356..39391
/rpt_family="AT_rich"
repeat_region complement(40008..40033)
/rpt_family="AT_rich"
repeat_region complement(42494..42754)
/rpt_family="AluSx"
repeat_region complement(43281..43318)
/rpt_family="(CA)n"
repeat_region 43318..43353
/rpt_family="(GA)n"
repeat_region complement(43880..44181)
/rpt_family="AluSx"
repeat_region 44713..44831
/rpt_family="(TAGAn)n"
repeat_region 45202..45304
/rpt_family="AluSx"
repeat_region 45540..45609
/rpt_family="L1PA5"
repeat_region 46081..46372
/rpt_family="AluY"
repeat_region complement(47354..47468)
/rpt_family="LINE2"
repeat_region 50273..50296

```

...: remainder of annotations omitted.

Query Match 5.1%; Score 94; DB 22; Length 148750;

Best Local Similarity 76.1%; Pred. No. 2.13e-54;

Matches 137; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

```

Db 114946 CTGAACACGGTGTGTGCTGCTGGTGCACCCGAGAGATGTCCTCTGCTGAGCTCC 115005
|||||  |||  |||||||||  |||||||||  |||||||||  |||
Cp 1369 CTGAGATTGAGATTCACCTGCTGGTGCATCGGATGAGTAGTGTCTCTGCTCAACTCT 1310
|||||  |||  |||||||||  |||||||||  |||||||||  |||
Db 115006 TTGAGTTGGCGGCCCCACAGTGTGCACGTGACCTGACCTGACCCCGAGATATCCAGG 115065
|||||  |||  |||||||||  |||||||||  |||||||||  |||
Cp 1309 TTGAGCTTAGCTGCTCCACATAGGTACAGACGGATCCCTCTAGAGTGTCTGG 1250
|||||  |||  |||||||||  |||||||||  |||||||||  |||
Db 115066 ATAGTCTTTCCACATCTCTCTTGTAGGAGACTTCACAGTCTTACCTCAGAGCTCTG 115125
|||||  |||  |||||||||  |||||||||  |||||||||  |||
Cp 1249 ATGGTATGTTCCACATCTCTTTAAAGGACTTCACATGTCCTTCCCTCTGAGGCTCTG 1190
|||||  |||  |||||||||  |||||||||  |||||||||  |||

```

```

RESULT 10
LOCUS HSCU177E8 40822 bp DNA PRI 27-JUL-1996
DEFINITION Human DNA sequence from cosmid cu177E8, between markers DXS366 and
DXS87 on chromosome X.
ACCESSION Z68694
NID 91160468
KEYWORDS X.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

REFERENCE
AUTHORS Williamson,H. and Wilkinson,P.
TITLE Direct Submision
JOURNAL Submitted (16-JAN-1995) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RQ, England. E-mail enquires: humpdes@sanger.ac.uk

```

```

COMMENT
IMPORTANT: This sequence is not the entire insert of clone
CU177E8. It may be shorter because we only sequence overlapping
sections once, or longer because we arrange for a small overlap
between neighbouring submissions.
The true left end of clone CU177E8 is at 1 in this sequence. The
true right end of clone CU177E8 is at 40822.
The true right end of clone V693A8 is at 34707.
CU177E8 is from the human chromosome X-specific cosmid library.

```

FEATURES

Source

```

1..40822
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="X"
/clone="CU177E8"

```

repeat_region

```
2..82
/partial
/note="Alu repeat: matches 210..291 of consensus"

```

repeat_region

```
88..119
/note="8 copies of 4 mer 100 & conserved"

```

repeat_region

```
88..123
/note="6 copies of 6 mer 94 & conserved"

```

repeat_region

```
88..121
/note="17 copies of 2 mer 100 & conserved"

```

repeat_region

```
304..383
/note="11 element fragment"

```

repeat_region

```
1367..1420
/note="3 copies of 18 mer 91 & conserved"

```

repeat_region

```
1367..1414
/note="8 copies of 6 mer 100 & conserved"

```

repeat_region

```
1367..1414
/note="12 copies of 4 mer 100 & conserved"

```

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

[illegible]

Query Match	Similarity	72.1%	Score 73	DB 22	Length 207
Matches 119	Conservative	0	Pred. No. 2,16e-36	Mismatches 46	Indels 0
				Gaps 0	
Db	28	TTTAGACCTGGAGCAGATTTTGTCATGCTGGAGAGAAATTTTCGGGTCATACGAGTG	87		
Qy	1028	TTTGGGGCAGAGAGCTGACTTCGTATGCTGGGTGCATGCTGGGACACAGTAGTC	1087		
Db	88	TGCTGGAGAGTGTTTGAGAGACGACGGAACTCTTTCACGGATAGCTC	147		
Qy	1088	AGGTGGTAGGCTCATCGAGAGGGAATGGCAGAGATCAAGCTCTTCATGAGATAGTTC	1147		
Db	148	TGACACCGCCATGAAACAGACCGCAGAGGAGATTCTAGTACAG	192		
Qy	1148	TGAATGGCCATGAAGAAGTATGCTGGGGGGCTGGCTAGTACAG	1192		

RESULT	LOCUS	DEFINITION	ACCESSION	NID	KEYWORDS	SOURCE	ORGANISM
13	EEOAMPHSM	4791 bp DNA	BCF	12-JAN-1995			
	Escherichia coli	ampd gene; guinolate phosphoribosyltransferase (nadc) gene; prepilin-like peptidase dependent protein (ppdp) gene; hnpB, hopc genes; GMP reductase (guac) gene.	L28105				
	9456039	GMP reductase: ampd gene; guac gene; hnpB gene; hopc gene; nadc gene; ppdp gene; prepilin-like peptidase dependent protein; guinolate phosphoribosyltransferase.					
	Escherichia coli	(strain K-12) DNA.					
	Escherichia coli	Eubacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.					
	REFERENCE	1 (sites)					
	AUTHORS	Andrews,S.C. and Guest J.R.					
	TITLE	Nucleotide sequence of the gene encoding the GMP reductase of Escherichia coli K12					
	JOURNAL	Biochem. J. 255 (1), 35-43 (1988)					
	MEDLINE	89061679					
	REFERENCE	2 (sites)					
	AUTHORS	Lindquist,S., Galleni,M., Lindberg,F. and Normark,S.					
	TITLE	Stannaling proteins in enterobacterial Ampc beta-lactamase regulation					
	JOURNAL	Mol. Microbiol. 3 (8), 1091-1102 (1989)					
	MEDLINE	90113890					
	REFERENCE	3 (sites)					
	AUTHORS	Hobbs,M. and Mattick,J.S.					
	TITLE	Common components in the assembly of type 4 fimbriae, DNA transfer systems, filamentous phage and protein-secretion apparatus: a general system for the formation of surface-associated protein complexes					
	JOURNAL	Mol. Microbiol. 10, 233-243 (1993)					
	MEDLINE	95020523					
	REFERENCE	4 (bases 1 to 4791)					
	AUTHORS	Whitchurch,C.B. and Mattick,J.S.					
	TITLE	Escherichia coli contains a set of genes homologous to those involved in protein secretion. DNA uptake and the assembly of type-4 fimbriae in other bacteria					
	JOURNAL	Gene 150 (1), 9-15 (1994)					
	MEDLINE	95047556					
	FEATURES	Location/Qualifiers					
	Source	1..4791					
	gene	/organism="Escherichia coli"					
	CDS	/strain="K-12"					
		/db_xref="taxon:562"					
		complement(1..290)					
		/gene="ampd"					
		complement(<1..290)					
		/gene="ampd"					
		/catation={2}					
		/codon_start=1					
		/db_xref="PID:g456040"					
		/transl_table=11					
		/translation="MTLEQGMVGARVPSPHDCRPDDETPTLLVYHNISLPGEFG"					

Query Match	Best Local Similarity	3.8%	Score 70:	DB 13:	Length 4791:
Matches 188:	Conservative	0:	Mismatches 118:	Indels 0:	Gaps 0:
<pre> /citation=[1] /codon_start=3 /product="GMP reductase" /db_xref="PID:g456045" /transl_table=11 /translation="MLAGHEGSGRIYEENGKFMLEFGMSSESAMKRVHGVAEYRA AEKRYKLPARGVENTANDIDGLRSACTYVAGSRLEKLTKRTTPIRVQDENRIRIN NL" </pre>					
Db	4486	TGAAGATGCGGTTTCTTCTTGTTCTCGACACAGCAATAAAGTGTCGCTGTCAGCTCT	4545		
Cp	1368	TGAAGATGATGATTACCTGCGTGGGTGACTGGAATGAAGATGTTCTCTGTCACACTTT	1309		
Db	4546	TCAGCGCTGAAGCCCCCAACGATATGTACAGCTGAACGACGCGCCCAAAATATCTGCG	4605		
Cp	1308	TGACCTTAGCTGCTGCCACATAGGTACACGTAGAGCGGATCCCTCTAGATGTCTCGGA	1249		
Db	4606	CGGATTTTCAACCGGGGCTTCGACGCGGACCTTAAGCGTTTACCTTCTGCTCGCAT	4665		
Cp	1248	TGATATGTTCCACATCTCTCTTTAAAGAACTTCCACTGCTTCTCCCTGAGGCTCTGT	1189		
Db	4666	ATTTCCGACACGCCCAACGATGAGCTTTCATCGACAGCTCGAGCTCATGCGGTAGACA	4725		
Cp	1188	ACTGACGACGCCGCCCAACGATGAGCTTTCATCGAGGCTATTTCAAGAACTATTCATAGAA	1129		
Db	4726	GCATAAATTTCTGCGCGCTTCTCTCAACGATGACGCGCGCTCTCTTCTGTCGCGCCA	4785		
Cp	1128	GCTTGATCTTCTGCAATCCCTCTGATGAGCTCACCACCTGACTACTGTGCCAGCA	1069		
Db	4786	GCATGC 4791			
Cp	1068	GCATGC 1063			
RESULT 14	LOCUS	LOCUS	LOCUS	PRI	10-JUL-1992
DEFINITION	2 of 9).	7303552	159 bp	DNA	guanosine monophosphate reductase [human, Genomic, 159 nt, segment
ACCESSION	73043	73043			
NID	9241125	9241125			
KEYWORDS	2 of 9				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;				
AUTHORS	Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
TITLE	1 (bases 1 to 159)				
JOURNAL	Kondoh,T., Kanno,H., Chang,L. and Yoshida,A.				
MEDLINE	Genomic structure and expression of human guanosine monophosphate				
REMARK	reductase				
	Hum. Genet. 88 (2), 219-224 (1991)				
	92098099				
FEATURES	GenBank staff at the National Library of Medicine created this				
	entry [NCBI g1b3673043] from the original journal article.				
	This sequence comes from Figure 5.				
	Location/Qualifiers				
	1..159				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
BASE COUNT	35 a	39 c	41 g	44 t	
ORIGIN					
Query Match	Best Local Similarity	3.6%	Score 66:	DB 22:	Length 159:
Matches 86:	Conservative	0:	Mismatches 20:	Indels 0:	Gaps 0:
Db	26	AGGTGATCTTGAACGACCTTCACGTTTCGAATTTCAAAAGACGCTACGAGGATTC	85		

This Page Blank (uspto)


```
DT 31-MAY-1994 (first entry)
DE Oligonucleotide probe MK14-A
KM Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW ss.
OS Synthetic.
PN EP-571911-A.
PD 01-DEC-1993.
PF 24-MAY-1993; 108325.
PR 26-MAY-1992; US-889651.
PA (BECT ) BECTON DICKINSON CO.
PI Shank DD, Spears PA;
DR WPI: 93-378844/8.
PT New oligo:nucleotide probes specific for Mycobacteria - used for
PT detection and amplification of Mycobacteria nucleic acid in
PT samples
PS Claim 3; Page 14; 23pp. English.
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC (Q51735). It hybridized to all spp. of mycobacteria tested, but
CC cross reacted to a few non-mycobacterial spp. The probe may
CC be useful as an initial screen for mycobacterial infection.
CC See also Q51735-45 and Q51747-59.
SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 2.4%; Score 44; DB 9; Length 91;
Best Local Similarity 0.0%; Pred. No. 4,42e-13;
Matches 0; Conservative 44; Mismatches 0; Indels 0; Gaps 0;

Db 13 vhsyyvvhvshhshvhhvhhvshvvvhhvhhvhhvhhv 56
CP 1036 GCCCCAAAGCCTTGCCACATCCCGACAGACGTCGACACCTCC 993

RESULT 3
ID N81164 standard; DNA; 204 BP.
AC N81164;
DT 08-NOV-1990 (first entry)
DE Base substituted E.coli beta-galactosidase alpha-fragment.
KM E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS Escherichia coli.
FH Key Location/Qualifiers
FT misc_feature 19..69
FT /tag= a
FT /function= multiple cloning site
FT primer_bind 187..204
FT /*tag= b

EP-285123-A.
PD 05-MAY-1988.
PF 30-MAR-1988; 105163.
PR 03-APR-1987; US-034819.
PA (SUSO) SUOMEN SOKERI OY.
PI Lehtovaara P, Knowles J, Kotivula A, Bamford J, Reinikainen T;
DR WPI: 88-279927/40.
PT Introducing random point mutations into nucleic acids -
PT by prepn of single stranded template, annealing a primer, elongation,
PT misincorporation, completion of molecules and screening.
PS Disclosure; p; English.
CC Random point mutations were introduced into the alpha fragment of
CC E.coli beta-galactosidase. The wild type sequence was obtained as a
CC single stranded template and an oligonucleotide was hybridised to
CC it to generate a popn of DNA molecules which terminate at all
CC possible nucleotide positions within a specified region. The
CC variable 3' ends generated in this way are used as primers for
CC reverse transcriptase. Nucleotides are misincorporated by the
CC transcriptase and the molecules are completed to forms that can be
CC amplified and then expressed in a suitable host-vector system.
CC The sequence covers all 176 diff base substitutions, most of which
CC occurred singularly in any given mutant.
CC See also P80575.
SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 2.4%; Score 45; DB 1; Length 204;
Best Local Similarity 8.7%; Pred. No. 9.73e-14;
Matches 8; Conservative 56; Mismatches 27; Indels 1; Gaps 1;
```

```
Db 93 hvirmbuvrvdyrdsaaawcyrrsvkydcymachdhvypbbvrvnvnhnn 152
CP 101 CTGGGAAGGTTGTGACAGACCTTGAGGGTGCAAT-CTTACCCCGCTACACTCTTC 43

Db 153 ccdbnhvchvbnhbnhvnwayrvhdaardv 184
CP 42 TTTACACCCCTGCAGAGCATGGGCGCAGAG 11

RESULT 4
ID N81164 standard; DNA; 204 BP.
AC N81164;
DT 08-NOV-1990 (first entry)
DE Base substituted E.coli beta-galactosidase alpha-fragment.
KM E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS Escherichia coli.
FH Key Location/Qualifiers
FT misc_feature 19..69
FT /tag= a
FT /function= multiple cloning site
FT primer_bind 187..204
FT /*tag= b

EP-285123-A.
PD 05-MAY-1988.
PF 30-MAR-1988; 105163.
PR 03-APR-1987; US-034819.
PA (SUSO) SUOMEN SOKERI OY.
PI Lehtovaara P, Knowles J, Kotivula A, Bamford J, Reinikainen T;
DR WPI: 88-279927/40.
PT Introducing random point mutations into nucleic acids -
PT by prepn of single stranded template, annealing a primer, elongation,
PT misincorporation, completion of molecules and screening.
PS Disclosure; p; English.
CC Random point mutations were introduced into the alpha fragment of
CC E.coli beta-galactosidase. The wild type sequence was obtained as a
CC single stranded template and an oligonucleotide was hybridised to
CC it to generate a popn of DNA molecules which terminate at all
CC possible nucleotide positions within a specified region. The
CC variable 3' ends generated in this way are used as primers for
CC reverse transcriptase. Nucleotides are misincorporated by the
CC transcriptase and the molecules are completed to forms that can be
CC amplified and then expressed in a suitable host-vector system.
CC The sequence covers all 176 diff base substitutions, most of which
CC occurred singularly in any given mutant.
CC See also P80575.
SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 2.3%; Score 42; DB 1; Length 204;
Best Local Similarity 10.2%; Pred. No. 8.90e-12;
Matches 11; Conservative 57; Mismatches 40; Indels 0; Gaps 0;

Db 78 dchvccgymrtctthvrmvbnvrvdyrdsaaawcyrrsvkydcymachdhv 137
QY 1270 TCCGCTACACTGTGACATGATGGAGACGTAAGCTCAAGATTGACGAGAACTA 1329

Db 138 ybbvbnvnnhnnccbnhchvbnhbnhvnwayrvhdaardv 185
QY 1330 CCTTCATCCGAGTCACCGACGAGTGAATCAATCTCAGTAGGCGT 1377

RESULT 5
ID Q06261 standard; DNA; 1829 BP.
AC Q06261;
DT 28-JAN-1991 (first entry)
DE Sequence encoding IMP dehydrogenase gene.
KW Inosine; guanosine; 5' inosinic acid; 5' guanylic acid; ds.
OS Bacillus sp.
PN EP-393969-A.
PD 24-OCT-1990.
PF 17-APR-1990; 304080.
PR 19-APR-1989; JP-101084.
PA (TAKE ) TAKEDA CHEMICAL IND KK.
```

PI Miyagawa K, Kanzaki N, Hasegawa T;
DR WPI; 90-322440/43.
PT Introduction of different expressible promoter into DNA - used
PS for improved prodn. of inosine and/or guanosine
PS disclosure: Fig 2: 21pp; English.
CC Sequence may be incorporated into a plasmid vector under the control
CC of a foreign expressible promoter, and used to transform a Bacillus
CC bacteria. New promoter may be used to express the gene at higher or
CC lower levels in pure form.
CC The products inosine and/or guanosine are the raw materials for the
CC production of 5'-IMP and 5'-GMP, flavour components of
CC "matsubushi" and "shitake mushroom".
SQ Sequence 1829 BP; 562 A; 359 C; 420 G; 488 T;
Query Match 2.3%; Score 43; DB 1; Length 1829;
Best Local Similarity 63.5%; Pred. No. 1,99e-12;
Matches 101; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
Db 1013 cattatgcggaagcgtgcaacagctgaagcgacaagagcgttatcgagctgagc 1072
QY 794 CATCATGCGAGGAGATGTGTACAGAGAGATGTAGAAAGCTCATCTTCTGGGCG 853
Db 1073 agagcgtgcgaagtgaatgagggcctggtcaattgtactacaagctgtgtagccgg 1132
QY 854 TGACATCATCAAAATGGGAAATTGGCCAGCTCTGTGTACTCTCGAAGAAACTGG 913
Db 1133 ggtggtctccgcaattacagcatttaattgttc 1171
QY 914 AGTGGGATCTCACAGCTCAGCGCAGTATGAGATGTC 952
RESULT 6
ID Q70469 standard; DNA; 114 BP.
AC Q70469;
DE 07-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSAR peptide library.
KW TSAR: totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc-feature 55..60
FT /*tag- a
FT /note- "this sequence represents 'Z'; Z can be a
FT sequence of 6,9 or 12 nucleotides (see
FT comments)"
PN MO9418318-A.
PD 18-AUG-1994.
PE 01-FEB-1994; U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI; 94-279739/34.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PS Disclosure; Page 35; 25pp; English.
CC Q70469 is a generic DNA sequence used to generate random TSAR peptide
CC library. This generic formula can be represented as follows: X(TGC)(NNB)10-
CC (TGC)(NNB)6Z(NNB)2(TGC)(NNB)14(TGC)Y. X and Y are flanking restriction
CC sites (X is not the same as Y) that are not specified further. This
CC sequence generates peptides that are cloverleaf in structure. Other
CC generic sequences are shown in Q70465-68. Other specific peptides
CC generated by these generic sequences are shown in R65150-54. TSARs are
CC concatenated heterofunctional proteins or peptides, comprising at least
CC two functional regions - a binding domain with affinity for a ligand and
CC a second effector peptide portion that is chemically or biologically
CC active. They may further comprise a linker peptide between the 2 domains.
CC The oligonucleotides are also designed so that the expressed peptide
CC contains 2 or 4 cysteine residues positioned in, or flanking, the
CC unpredicted or variant residues. These residues confer some degree of

CC conformational rigidity to the peptides. The TSARs or compsns. comprising
CC a TSAR binding domain can be used in vivo to deliver a chemically or
CC biologically active moiety, eg. metal ion, radioisotope, peptide, toxin
CC or enzyme, to the specific target or on the cell. They can also replace
CC the function of macromolecules, eg. monoclonal or polyclonal antibodies
CC and therefore circumvent the need for complex methods of hybridoma
CC formation or in vivo antibody production. The TSARs are easily
CC characterised and have designed activity allowing direct and rapid
CC detection in a screening process.
SQ Sequence 114 BP; 0 A; 4 C; 4 G; 4 T;
Query Match 2.0%; Score 36; DB 12; Length 114;
Best Local Similarity 5.6%; Pred. No. 5,70e-08;
Matches 6; Conservative 32; Mismatches 70; Indels 0; Gaps 0;
Db 6 bnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnb 65
QY 482 CATCATGCTGCCAATATGGAATCTGTGGCACCCTTGACATGCGCAAGCTTCTGTAA 541
Db 66 bbgcnbhb 113
QY 542 GTTCTCTCTTCACTGCTGCTCCATAGCACTATAGCCTGCTCAGTG 589
RESULT 7
ID Q70465 standard; DNA; 114 BP.
AC Q70465;
DE 05-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSAR peptide library.
KW TSAR: totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc-feature 55..60
FT /*tag- a
FT /note- "this sequence represents 'Z'; Z can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"
PN MO9418318-A.
PD 18-AUG-1994.
PE 01-FEB-1994; U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI; 94-279739/34.
DR P-PSDB; R65150 and R65151.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PS Disclosure; Page 35; 25pp; English.
CC Q70465 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides. This generic formula can also be
CC represented as follows: X(NNB)6(TGC)(NNB)11Z(NNB)14(TGC)(NNB)3Y. X
CC and Y are flanking restriction sites (X is not the same as Y) that are
CC not specified further. Other generic sequences are shown in Q70466-68.
CC Other specific peptides generated by these generic sequences are shown in
CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC comprising at least two functional regions - a binding domain with
CC affinity for a ligand and a second effector peptide portion that is
CC chemically or biologically active. They may further comprise a linker
CC peptide between the 2 domains. The oligonucleotides are also designed so
CC that the expressed peptide contains 2 or 4 cysteine residues positioned
CC in, or flanking, the unpredicted or variant residues. These residues
CC confer some degree of conformational rigidity to the peptides. The TSARs
CC or compsns. comprising a TSAR binding domain can be used in vivo to
CC deliver a chemically or biologically active moiety, eg. metal ion,
CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need
CC for complex methods of hybridoma formation or in vivo antibody

production The TSARs are easily characterised and have designed activity allowing direct and rapid detection in a screening process.

SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;

Query Match 1.8%; Score 34; DB 12; Length 114;
Best Local Similarity 2.7%; Pred. No. 9.63e-07;
Matches 3; Conservative 34; Mismatches 75; Indels 0; Gaps 0;

Ddb 3 bnbnnbnbnbnbnbnbtgcnnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnnnn 62
OY 476 GGTTCCCATGTCATGTGCCAATAATGGACTAGTGGACACTTGAGATGGCAAGCTTC 535
Ddb 63 bnbnnbn 114
OY 536 CTGTAACTTCTCCTCTCACCTGCTGCATTAGAACAATTAAGCCTCGTTCAg 587

RESULT 8
ID O70467 standard; DNA; 114 BP.
AC O70467;
DT 05-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSAR peptide library.
KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KM effector domain; concatenated heterofunctional protein; linker;
KN direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_feature 55..60 /*tag= a
FT /note= "this sequence represents 'Z'; Z can be a
FT sequence of 6, 9 or 12 nucleotides (see comments)"
FN WC9418318-A.
PD 18-AUG-1994.
PR 01-FEB-1994; U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
PT WPI: 94-279739/34.
DR P-PDSB; R65153.
PT Identifying proteins or peptide(s) which bind a ligand - by
FT screening a recombinant vector library expressing fusion proteins
PS comprising a binding domain and an effector domain
PN Disclosure: Page 35; 255pp; English.
CC O70467 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides. This generic formula can also be
CC represented as follows: X(NNB)16(TGC)(NNB)12(GCN)(NNB)1X. X
CC and Y are flanking restriction sites (X is not the same as Y) that are
CC not specified further. Other generic sequences are shown in O70466-68.
CC Other specific peptides generated by these generic sequences are shown in
CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC comprising at least two functional regions - a binding domain with
CC affinity for a ligand and a second effector peptide portion that is
CC chemically or biologically active. They may further comprise a linker
CC peptide between the 2 domains. The oligonucleotides are also designed so
CC that the expressed peptide contains 2 or 4 cysteine residues positioned
CC in, or flanking, the unpredicted or variant residues. These residues
CC confer some degree of conformational rigidity to the peptides. The TSARs
CC or comprints, comprising a TSAR binding domain can be used in vivo to
CC deliver a chemically or biologically active moiety, eg. metal ion,
CC radiolabeled peptide, toxin or enzyme, to the specific target or on the
CC cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need for
CC complex methods of hybridoma formation or in vivo antibody production.
CC The TSARs are easily characterised and have designed actively allowing
CC direct and rapid detection in a screening process.
SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;

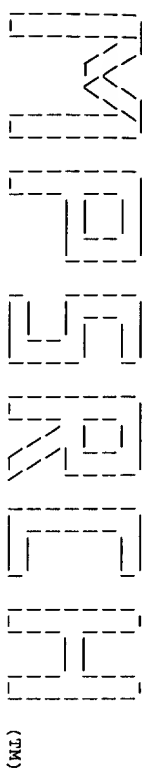
Query Match 1.8%; Score 34; DB 12; Length 114;
Best Local Similarity 2.7%; Pred. NO. 9.63e-07;
Matches 3; Conservative 34; Mismatches 75; Indels 0; Gaps 0;

[illegible]

AC 070465; (first entry)
 DT 05-APR-1995
 DE Generic DNA sequence to generate a random TSAR peptide library.
 DE TSAR: totally synthetic affinity reagent; synthetic; binding domain;
 KM effector domain: concatenated heterofunctional protein; linker;
 KW direct; rapid; detection; screening; treatment; generic; ss.
 OS Synthetic.
 FH Key
 FT misc_feature
 FT 55..60
 FT /tag- a
 FT /note- "this sequence represents 'Z'; Z can be a
 FT sequence of 6, 9 or 12 nucleotides (see
 FT comments)"
 PN WO9418318-A.
 PD 18-AUG-1994.
 PF 01-FEB-1994; US-013416.
 PR 01-FEB-1993; US-013416.
 PR 30-DEC-1993; US-176500.
 PR 31-JAN-1994; US-189331.
 PA (UYN-) UNIV NORTH CAROLINA.
 PI Fowlkes DM, Kay BK;
 DR FPI: 94-279739/34.
 DR P-PSDB: R65150 and R65151.
 PT Identifying proteins or peptide(s) which bind a ligand - by
 PT screening a recombinant vector library expressing fusion proteins
 PT comprising a binding domain and an effector domain
 PS Disclosure; Page 35; 255PP; English.
 CC 070465 is a generic DNA sequence used to generate random TSAR (Totally
 CC Synthetic Affinity Reagents) peptides. This generic formula can also be
 CC represented as follows: X(NNB)6(TGC)(NNB)11(TGC)(NNB)3Y. X
 CC and Y are flanking restriction sites (X is not the same as Y) that are
 CC not specified further. Other generic sequences are shown in 070466-68.
 CC Other specific peptides generated by these generic sequences are shown in
 CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
 CC comprising at least two functional regions - a binding domain with
 CC affinity for a ligand and a second effector peptide portion that is
 CC chemically or biologically active. They may further comprise a linker
 CC peptide between the 2 domains. The oligonucleotides are also designed so
 CC that the expressed peptide contains 2 or 4 cysteine residues positioned
 CC in, or flanking, the unpredicted or variant residues. These residues
 CC confer some degree of conformational rigidity to the peptides. The TSARs
 CC or compns. comprising a TSAR binding domain can be used in vivo to
 CC deliver a chemically or biologically active moiety, eg. metal ion,
 CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
 CC cell. They can also replace the function of macromolecules, eg.
 CC monoclonal or polyclonal antibodies and therefore circumvent the need
 CC for complex methods of hybridoma formation or in vivo antibody
 CC production. The TSARs are easily characterised and have designed
 CC activity allowing direct and rapid detection in a screening process.
 SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;
 Query Match 1.8%; Score 34; DB 12; Length 114;
 Best Local Similarity 3.6%; Pred. NO. 9.63e-07;
 Matches 4; Conservative 33; Mismatches 75; Indels 0; Gaps 0;
 Db 3 bnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnnnnnn 62
 Cp 903 TCCGAGTAGTACACACAGAGCGCTGCCCATTTCCATTGATATGTCACCCACAGAA 844
 Db 63 bnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnnb 114
 Cp 843 GGATGACCTCTTACCATCTCTCCTGTACACATTCCTCCATGATGTGT 792
 RESULT 11
 ID 070468 standard; DNA; 114 BP.
 AC 070468;
 DT 05-APR-1995 (first entry)
 DE Generic DNA sequence to generate a random TSAR peptide library.
 DE TSAR: totally synthetic affinity reagent; synthetic; binding domain;
 KM effector domain: concatenated heterofunctional protein; linker;
 KW direct; rapid; detection; screening; treatment; generic; ss.
 OS Synthetic.
 FH Key
 FT misc_feature
 FT 55..60
 FT /tag- a
 FT /note- "this sequence represents 'Z'; Z can be a
 FT sequence of 6, 9 or 12 nucleotides (see
 FT comments)"
 PN WO9418318-A.

FH Key
 FT misc_feature
 FT 55..60
 FT /tag- a
 FT /note- "this sequence represents 'Z'; Z can be a
 FT sequence of 6, 9 or 12 nucleotides (see
 FT comments)"
 PN WO9418318-A.
 PD 18-AUG-1994.
 PF 01-FEB-1994; US-013416.
 PR 01-FEB-1993; US-013416.
 PR 30-DEC-1993; US-176500.
 PR 31-JAN-1994; US-189331.
 PA (UYN-) UNIV NORTH CAROLINA.
 PI Fowlkes DM, Kay BK;
 DR FPI: 94-279739/34.
 DR P-PSDB: R65154.
 PT Identifying proteins or peptide(s) which bind a ligand - by
 PT screening a recombinant vector library expressing fusion proteins
 PT comprising a binding domain and an effector domain
 PS Disclosure; Page 35; 255PP; English.
 CC 070468 is a generic DNA sequence used to generate random TSAR (Totally
 CC Synthetic Affinity Reagents) peptides. This generic formula can also be
 CC represented as follows: X(NNB)11(TGC)(NNB)62(TGC)(NNB)10Y. X
 CC and Y are flanking restriction sites (X is not the same as Y) that are
 CC not specified further. Other generic sequences are shown in 070466-68.
 CC Other specific peptides generated by these generic sequences are shown in
 CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
 CC comprising at least two functional regions - a binding domain with
 CC affinity for a ligand and a second effector peptide portion that is
 CC chemically or biologically active. They may further comprise a linker
 CC peptide between the 2 domains. The oligonucleotides are also designed so
 CC that the expressed peptide contains 2 or 4 cysteine residues positioned
 CC in, or flanking, the unpredicted or variant residues. These residues
 CC confer some degree of conformational rigidity to the peptides. The TSARs
 CC or compns. comprising a TSAR binding domain can be used in vivo to
 CC deliver a chemically or biologically active moiety, eg. metal ion,
 CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
 CC cell. They can also replace the function of macromolecules, eg.
 CC monoclonal or polyclonal antibodies and therefore circumvent the need
 CC for complex methods of hybridoma formation or in vivo antibody
 CC production. The TSARs are easily characterised and have designed activity
 CC allowing direct and rapid detection in a screening process.
 SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;
 Query Match 1.8%; Score 33; DB 12; Length 114;
 Best Local Similarity 2.8%; Pred. NO. 3.88e-06;
 Matches 3; Conservative 33; Mismatches 73; Indels 0; Gaps 0;
 Db 3 bnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnnnnnn 62
 Cp 1652 GGATGTTGTTGGTCCAGATTCCTATTGATTTCTGTCATCATTTCCACACTG 1593
 Db 63 bnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnbdbnnb 111
 Cp 1592 TTTCCTCTCTGCTCTCTCCAGTGTGCTGCTGCGCATTTGTGTGT 1544
 RESULT 12
 ID 070467 standard; DNA; 114 BP.
 AC 070467;
 DT 05-APR-1995 (first entry)
 DE Generic DNA sequence to generate a random TSAR peptide library.
 DE TSAR: totally synthetic affinity reagent; synthetic; binding domain;
 KM effector domain: concatenated heterofunctional protein; linker;
 KW direct; rapid; detection; screening; treatment; generic; ss.
 OS Synthetic.
 FH Key
 FT misc_feature
 FT 55..60
 FT /tag- a
 FT /note- "this sequence represents 'Z'; Z can be a
 FT sequence of 6, 9 or 12 nucleotides (see
 FT comments)"
 PN WO9418318-A.

This Page Blank (uspto)



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPearch_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Wed Jun 17 12:07:46 1998; MasPar time 1802.52 Seconds

Tabular output not generated. 1365.733 Million cell updates/sec

Title: >US-08-774-169-2
(1-1843) from US08774169.seq

Description: 8:gb_est14 9:gb_est15 10:gb_est16 11:gb_est17

Perfect Score: 1843 1 CTAATCAGCCTCTGCCCC.....AGAGAGCTCATTGACTGT 1843

N.A. Sequence: Comp: GATTGATCGAGACAGCGGG.....TCCTCTCGAGTACAGACA

Scoring table: TABLE default
Gap 6

Mmatch STD : Dbase 0; Query 0

Searched: 1759237 seqs, 667866413 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: emb1-est54
1:em_est1 2:em_est3

Database: genbank-est106

3:gb_est14 4:gb_est10 5:gb_est11 6:gb_est12 7:gb_est13
8:gb_est14 9:gb_est15 10:gb_est16 11:gb_est17
12:gb_est18 13:gb_est19 14:gb_est2 15:gb_est10
16:gb_est12 17:gb_est22 18:gb_est3 19:gb_est5
21:gb_est6 22:gb_est7 23:gb_est8 24:gb_est9 25:gb_est5
26:gb_est5

Statistics: Mean 11.691; Variance 1.920; scale 6.088

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	604	32.8	635	8	AA507892	nh75c02.s1 NCI_CGAP.Br	0.00e+00
2	549	30.8	646	22	W60937	zdz7e06.r1 Soares feta	0.00e+00
3	549	29.8	632	22	W65156	zcs8e05.r1 Soares para	0.00e+00
4	502	27.2	731	23	AA172963	ms58b04.r1 Life Tech m	0.00e+00
5	491	26.6	494	13	AA009421	ze82c03.r1 Soares feta	0.00e+00
6	489	26.5	501	19	N28887	ye63a11.r1 Homo sapien	0.00e+00
7	485	26.3	546	19	N30305	ye69h06.s1 Homo sapien	0.00e+00
8	464	25.2	472	13	AA778158	zfa5b09.s1 Soares feta	0.00e+00
9	465	25.2	474	19	N35234	yy21d05.s1 Homo sapien	0.00e+00
10	465	25.2	478	7	AA456771	zw27f02.r1 Soares ovar	0.00e+00
11	460	25.0	599	9	AA524317	ng33h05.s1 NCI_CGAP.Pr	0.00e+00
12	452	24.5	473	11	AA654123	nt10b06.s1 NCI_CGAP.Pr	0.00e+00
13	450	24.4	463	19	H99254	yx21h07.s1 Homo sapien	0.00e+00

RESULT	LOCUS	DEFINITION	ACCESSION	NID	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	AA507892	635 bp mRNA	EST	18-AUG-1997	nh75c02.s1 NCI_CGAP.Br.1 Homo sapiens cDNA clone IMAGE:964332	similar to gb:M24470 GMP REDUCTASE (HUMAN);, mRNA sequence.	human.	Human sapiens	Robert Strausberg, Ph.D.	Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov Tissue Procurement: Christopher Koskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: M. Bento Soares, Ph.D. DNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/dbp/image/image.html	Insert Length: 1257 Std Error: 0.00 Seq primer: -40m3 fwd. ET from Amersham High quality sequence stop: 448. Location/Qualifiers 1. 635 /organism="Homo sapiens" /note="Vector: pT73D-Pac (Pharmacia) with a modified	

ALIGNMENTS

RESULT	LOCUS	DEFINITION	ACCESSION	NID	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	AA507892	635 bp mRNA	EST	18-AUG-1997	nh75c02.s1 NCI_CGAP.Br.1 Homo sapiens cDNA clone IMAGE:964332	similar to gb:M24470 GMP REDUCTASE (HUMAN);, mRNA sequence.	human.	Human sapiens	Robert Strausberg, Ph.D.	Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov Tissue Procurement: Christopher Koskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: M. Bento Soares, Ph.D. DNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/dbp/image/image.html	Insert Length: 1257 Std Error: 0.00 Seq primer: -40m3 fwd. ET from Amersham High quality sequence stop: 448. Location/Qualifiers 1. 635 /organism="Homo sapiens" /note="Vector: pT73D-Pac (Pharmacia) with a modified	

polylinker: 1st strand cDNA was prepared from pooled bulk breast tumor tissue, and was then primed with a Not I - o1igo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is not normalized. (The normalized version of this library is NCI CGAP Br2.) Library was constructed by Bento Soares and M. Fatima Bonaldo."

/db_xref="taxon:9606"
/clone="IMAGE:964322"
/clone_lib="NCI CGAP Br1.1"
/sex="female, pooled"
/tissue_type="breast"
/lab_host="DH10B"
<1. >635

BASE COUNT 161 a 139 c 149 g 186 t
ORIGIN

Query Match 32.8%; Score 604; DB 8; Length 635;
Best Local Similarity 98.6%; Pred. NO. 0.00e+00;
Matches 623; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

Db 1 ACAGTCATGAGCTCTCTCTTATTTAGAGTATTCATAATATCAAAACACTTCAGAGCTTT 60
CP 1843 ACAGTCATGAGCTCTCTCTTATTTAGAGTATTCATAATATCAAAACACTTCAGAGCTTT 1784
Db 61 TTAGATATGTGAAGTCCCAATTAGTGAAGGGATTAAACAAACATGATTTTAAAAAGCCT 120
CP 1783 TTAGATATGTGAAGTCCCAATTAGTGAAGGGATTAAACAAACATGATTTTAAAAAGCCT 1724
Db 121 GCCCCTCTAGACCAAGGCGCATGTAAAGATTATGTGCAATCGCATCGCTTTCCCTTT 180
CP 1723 GCCCCTCTAGACCAAGGCGCATGTAAAGATTATGTGCAATCGCATCGCTTTCCCTTT 1664
Db 181 AATAATCTTCAGAGTGTGTGTGGTCCCAATTCCTTGTATTTCTTGCAATCT 240
CP 1663 AATAATCTTCAGAGTGTGTGTGGTCCCAATTCCTTGTATTTCTTGCAATCT 1604
Db 241 TTCACAGCTGTGTCTCTCTCTCTCCCAAGTGAAGTCCCTGGAGTTTGTGTGT 300
CP 1603 TTCACAGCTGTGTCTCTCTCTCTCCCAAGTGAAGTCCCTGGAGTTTGTGTGT 1544
Db 301 GCACATAGAGAGTACAGAGTACTGCGAGAGCCCTGAGAGTGAAGACAAAGAGAAATGA 360
CP 1543 GCACATAGAGAGTACAGAGTACTGCGAGAGCCCTGAGAGTGAAGACAAAGAGAAATGA 1484
Db 361 CAAATATATACAGAGCTGACAGTACCTGGATGGAGAGACCCCACTGGAGTGGCCCAT 420
CP 1483 CAAATATATACAGAGCTGACAGTACCTGGATGGAGAGACCCCACTGGAGTGGCCCAT 1424
Db 421 GGTAGAGTACTGTGCTTGGAGGGTGAACCTGCTCAAGTCTAGACAGCCCTCACTGAAG 480
CP 1423 GGTAGAGTACTGTGCTTGGAGGGTGAACCTGCTCAAGTCTAGACAGCCCTCACTGAAG 1364
Db 481 ATTGATTCACCTCTGGTGCATCTCGATGAAGTATCTTCTGCTCAACTTTTGAAC 540
CP 1363 ATTGATTCACCTCTGGTGCATCTCGATGAAGTATCTTCTGCTCAACTTTTGAAC 1304
Db 541 TTAGTGTCTCCACATAGTACAGTACAGAGGATCCATCTAGAGTCTCGATGAGTA 600
CP 1303 TTAGTGTCTCCACATAGTACAGTACAGAGGATCCATCTAGAGTCTCGATGAGTA 1244
Db 601 TGT-CCACATCTCGTTTGAAGAAG-CTTCCAC 630
CP 1243 TGTCCACATCTCGTTTGAAGAAG-CTTCCAC 1212

RESULT 2
LOCUS W60937 646 bp mRNA EST 15-OCT-1996
DEFINITION z427e06.r1 Soares fetal heart NBH19W Homo sapiens cDNA clone
341890 5' similar to gb:M24470 GMP REDUCTASE (HUMAN);, mRNA
sequence.
ACCESSION W60937

NIID 61367705
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1993)

COMMENT

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 810 Std Error: 0.00
Seq primer: mod.RegA+ET
High quality sequence stop: 400.

FEATURES
source location/Qualifiers
1. 646

/organism="Homo sapiens"
/note="Organ: heart; Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - o1igo(dT) primer [5' TGTACCAATCTGACGTGAGAGCGGCGCATGTATTTTGTGTGTGT 3']; double-stranded cDNA was size selected, ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M. Fatima Bonaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NBH19W."

/db_xref="taxon:9606"
/clone="341890"
/clone_lib="Soares fetal heart NBH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
<1. >646

BASE COUNT 183 a 150 c 161 g 148 t 4 others
ORIGIN

Query Match 30.8%; Score 568; DB 22; Length 646;
Best Local Similarity 97.4%; Pred. NO. 0.00e+00;
Matches 636; Conservative 0; Mismatches 6; Indels 11; Gaps 11;

Db 2 GAGCTCATCGAGAGGATGCGAAGATACAGCTCTTATGGAATGAGTCTGAATG 61
QY 1095 GAGCTCATCGAGAGGATGCGAAGATACAGCTCTTATGGAATGAGTCTGAATG 1154
QY 1155 GCCATGGAAGAAGTATGCTGGGGGCGTGTGAGTACAGAGCCCTCAGAGGGAAGACAGTG 1214
Db 122 GAATTCCTTTTAAAGAGATGTGGAACATACATCCGAGACATCCATAGAGAGGATCCGC 181
QY 1215 GAATTCCTTTTAAAGAGATGTGGAACATACATCCGAGACATCCATAGAGAGGATCCGC 1274
Db 182 TCTACGTGTAAGTGTGGAGAGAGCTTACCTCAAGAGTGTGAGAGAGAGTACCTTC 241
QY 1275 TCTACGTGTAAGTGTGGAGAGAGCTTACCTCAAGAGTGTGAGAGAGAGTACCTTC 1334
Db 242 ATCCGAGTCAACCGACGAGTGAATCCATCTTCAGTGAAGCGTGTAGACCTGAGCAAGT 301

QY	1335	ATCCGAGCACCACCGACAGAGTGAATTCATCTTCAGTAGGCGGTGCTAGACCTGAGCA-GT	1393
Db	302	TTTACCTCTCCCAAGGACGACCAAGTACTCTACCATATGGGGCATGCCAAGTTGGGCTCTCACC	361
QY	1394	TTCTACCTCTCCCAAGGACCA-GTACTCTACCATATGGGGCATGCCAAGT-GGGGTCTCTACC	1451
Db	362	CATCCACAGCTACTGAGCTGTATTTATTTTGTCATTTCTGTTGTCTGCTGTCACCTCGAAGG	421
QY	1452	CATCCACAGCTACTGAGCTGTATTTATTTTGTCATTTCTGTTGTCTGCTGTCACCTCGAAGG	1511
Db	422	CTCTGAGTAACTCTGTACTTCTCTATCTCAGACACACAAAATGCCCAAGCAGCTCAGTG	481
QY	1512	CTCTGAGTAACTCTGTACTTCTCTATCTCAGACACACAAAATGCCCAAGCAGCTCAGTG	1571
Db	482	GGGAGGAGCAAGGAGCAACAGCTGTGAGAAATGATGCAAGAA-TCAAATGGGAATC	540
QY	1572	GGGAGGAGCAAGGAGCAACAGCTGTGAGAAATGATGCAAGAAATCAATGGGAATC	1631
Db	541	TGGGGAGACCA-CACAACTCTCTGAAGATTATTTNAAAGAAA-GATGCTGATTG-TACTATA	597
QY	1632	TGGGGAGACCAACAACTCTCTGAAGATTATTTNAAAGAAAAGATGCTGATTGTTGATATA	1691
Db	598	A-TCTTTACATGGGCT-GGCTAGAG-AGGAGGAGCTTTAGA-TCATGTTT	646
QY	1692	AATCTTTACATGGGCTGTCTAGAGAGGAGGAGCTTTAGAAATCATGTTT	1744

RESULT	3	W56156	632 bp	mRNA	EST	11-OCT-1996
LOCUS		25588605.r1	Soares parathyroid tumor	NbHRA	Homo sapiens	CDNA clone
DEFINITION		326528 5', mRNA sequence.				
ACCESSION		W56156				
NID		q1358045				
KEYWORDS		EST.				
SOURCE		human.				
ORGANISM		Homo sapiens				
REFERENCE		Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Carnivora; Homidae; Homo.				
AUTHORS		1 (bases 1 to 632)				
		Hallier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Riklin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P., and Wilson, R.				
TITLE		The Washu-Merck EST Project				
JOURNAL		Unpublished (1995)				
COMMENT		Contact: Wilson RK Washu-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 1472 Std Error: 0.00 Seq primer: mob. REGA+ET High quality sequence stop: 404. Location/Qualifiers 1. 632 /organism="Homo sapiens" /note="Organ: parathyroid gland; Vector: pRT73D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - 5'- TGTTACCAATCTGAAGGGAGGCGGACCAATTTTTTTTTTTTTTTTTTTT T-3'", double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pRT733 vector (Pharmacia) Library went through one round of				

LOCUS	DEFINITION	Accession	NID	KEYWORDS	SOURCE
AA172963	731 bp mRNA	EST	23-DEC-1996		
ms58BD4.r1	Life Tech mouse embryo 13 5dpc 10666014 Mus musculus cDNA clone 615727 5' similar to gb:m24470 GMP REDUCTASE (HUMAN), mRNA sequence.	AA172863	91752032	EST.	house mouse.

ORGANISM	Mus musculus
REFERENCE	Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 731) Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Garske,J., Hall,L., Hillier,L., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theisberg,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE	The WashU-HHMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.edu This clone is available royalty-free through LNL, contact the IMAGE Consortium (info@image.lnl.gov) for further information. MG:376551 Seq primer: -28M13 rev1 from Amersham High quality sequence stop: 428. location/Qualifiers 1..731 /organism="Mus musculus" /strain="C57BL/6J" /note="Vector: PCMV-SPORT2; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. 13.5dpc embryos. PCMV-SPORT2 vector." /db_xref="taxon:10090" /clone="615721" /clone_1id="14fe Tech mouse embryo 13 5dpc 10666014" /dev_stage="13.5dpc embryos" /lab_host="DH10B" <1..>731
BASE COUNT	193 a 165 c 183 g 189 t 1 others
ORIGIN	nrna
Query Match	27.2%; Score 502; DB 23; Length 731; Best Local Similarity 89.9%; Pred. NO. 0.00e+00; Matches 570; Conservative 0; Mismatches 63; Indels 1; Gaps 1
Db	95 CATGCCATATTGATATGATGATGGAACCTTGACTTCAGAGATCTCTGTTGAGGCCAA 154
QY	335 CATGCCATATTGATGGAACGATGGAACCTTGACTTCAGAGATCTCTGTTGAGGCCAA 394
Db	155 ACGCAGCACTTTAAGTGTCCAGAGAGAGTACGTCACAGATCATCTTGTCCGAA 214
QY	395 ACGCAGTACCTTAAGTGTCCAGAGAGTACGTCACAGATCATCTTGTCCGAA 454
Db	215 CTCAAAGCAGATGTACAGTGGCATCCCTGTCATGCTGCCATATGATACCGTGGCAC 274
QY	455 CTCAAAGCAGATCATCTGCGGGTCCCATCATGTCGCCAATATGATATGATGAGGCAC 514
Db	275 TTTTGAATGGCTAGGGTCTCTGTGAAGTTCCTCGTACCTGCCATCCATAGACATA 334
QY	515 CTGTGAGTGGCCAAAGTTCTCTGTGAAGTTCCTCTCTCACTCTGTCCATAAGACATA 574
Db	335 CAGCATCATAGTGGCAAGAGTTTGGTCAGGCCAAGATCCTACTGTCTGAGTGTACGC 394
QY	575 TAGCTCGTTACGTGGCAAGAGTTTGGTCAGGCCAAGATCCTACTGTCTGAGCATCTGC 634
Db	395 TGCCAGCTCAGGCTCAGGCTCTGCTGACTTTGAGCAGCTGGAACAGATCCTGGAAGCTAT 454
QY	635 TGCCAGCTCAGGCAAGGCTCTGCTGACTTTGAGCAGCTGGAACAGATCCTGGAAGCTAT 694
Db	455 TCCCAATCAATATATATGCTCGGTGAGTGGCTAAGGCTACTCTGAACATTTTGTGA 514
QY	695 TCCCAAGTGAATATATGCTCGGTGAGTGGCAAGTGGCTACTCTGAACATTTTGTGA 754

	Db	515	ATTGTAAAGATGTTACGGAAAGCATTCCTCCCAACACACCATCATGCAGGAATGTGGT	574
	Oy	755	ATTGTAAAGATGTACCGAAGCAGCGCTTCCCCAGCACACATCATGCCAGGAATGTGGT	814
	Db	575	AACAGAGAAGATGGTAGAAGAGTTAAATCCCTTTCTGGGGCTACATCATCAAAGTGGGAAA	634
	Oy	815	AACAGAGAAGATGGTAGAAGAGCTCATCTTTCTGGGGCTACATCATCAAAGTGGGAAAT	874
	Db	635	TGTCCTGCTGCTCTGTCTGTACACTTCGAGAAAACTGAGTTNGTATC-ACAGCTCG	693
	Oy	875	TGGCGCAGGACTCTGTGTACTACTACGGAAGAAAACTGAGTGGGGTATCCACAGCTCAG	934
	Db	694	TGCAGTGATGCAGAGTTCAGATGCTGCTCAGGC	727
	Oy	935	CGCAGTGATGGAGTGTGCAGATGCTGCTCATGCG	968
LOCUS	5	AA009421	494 bp	mRNA EST 01-FEB-1997
DEFINITION		zee2c65.tl Soares fetal heart NBH19W Homo sapiens CDNA clone		
ACCESSION		AA009421		
NID		91470619		
KEYWORDS		EST.		
SOURCE		human.		
ORGANISM		Homo sapiens		
REFERENCE		Eukaryotes; mitochondriaal eukaryotes; Metazoa; Chordata; Vertebrates; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 494)		
AUTHORS		Hillier,L., Clark,N., Dubnue,T., Elliston,K., Hawkins,M., Holman,M., Hiltman,M., Kucaba,T., Le,M., Lennon,G., Maitra,M., Parsons,D., Riklin,L., Ronlfing,T., Tan,F., Trevaskis,E., Waterson,R., Williamson,A., Wohldmann,P. and Wilson.R.		
TITLE		Washington-Merck EST project		
JOURNAL		Unpublished (1995)		
COMMENT		Contact: Wilson RK WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information. Insert Length: 638 Std Error: 0.00 Seq primer: mob.REGA+ET.		
FEATURES		Location/Dualifiers		
SOURCE		1..494		
		/organism="Homo sapiens"		
		/note="Organ: heart; Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTACCATCATGTGAGTGGAGCGCGCCGATCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Felina Bonardo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NBH19W."		
		/db_xref="taxon:9606"		
		/clone_id="365480"		
		/clone_id="Soares fetal heart NBH19W"		
		/sex="unknown"		
		/dev_stage="19 weeks"		
		/lab_host="DH10B (ampicillin resistant)"		
		<1..>494		
BASE COUNT		140 a 126 c 117 g 111 t		
ORIGIN				

Query Match 26.6%; Score 491; DB 23; Length 494;
 Best Local Similarity 99.8%; Pred. No. 0.00e+00;
 Matches 492; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 2 GTGGAACATACCATCCGAGACATCTAGAGAGATCCGCTCTAGCTGTACTGTAGTGGGA 61
 |||||||
 QY 1236 GTGGAACATACCATCCGAGACATCTAGAGAGATCCGCTCTAGCTGTACTGTAGTGGGA 1235
 Db 62 GCACCTAAGCTCAAAAGAGTGGAGCAGAGAACTACCTTCATCCGAGTACCACGAGGTG 121
 |||||||
 QY 1296 GCACCTAAGCTCAAAAGAGTGGAGCAGAGAACTACCTTCATCCGAGTACCACGAGGTG 1355
 Db 122 AATCAATCTCATGTAGAGGCGCTGTAGACCTGAGAGAGTGTACCTTCCAGAGCAGCAT 181
 |||||||
 QY 1356 AATCAATCTCATGTAGAGGCGCTGTAGACCTGAGAGAGTGTACCTTCCAGAGCAGCAT 1415
 Db 182 ACTGTACCATGGGCGATCCCAAGTGGGGTCTCACCCTACCTCCAGTACTGAGCTGTETA 241
 |||||||
 QY 1416 ACTGTACCATGGGCGATCCCAAGTGGGGTCTCACCCTACCTCCAGTACTGAGCTGTETA 1475
 Db 242 TTACTTGTCTATTTCTCTGTCTCTCTACTCTCTGAGGCTCTGCAAGTACTGTACTTCT 301
 |||||||
 QY 1476 TTACTTGTCTATTTCTCTGTCTCTCTACTCTCTGAGGCTCTGCAAGTACTGTACTTCT 1535
 Db 302 CTATCTGACACACAAATGGCCCAAGGCACTCATCTGGGAGAGAAAGCAAGCAAGCAG 361
 |||||||
 QY 1536 CTATCTGACACACAAATGGCCCAAGGCACTCATCTGGGAGAGAAAGCAAGCAAGCAG 1595
 Db 362 TCTGAGAAATGATGACAGAAATCAATGGGAATCTGGGAGCCCAACACAACTCTCTGA 421
 |||||||
 QY 1596 TCTGAGAAATGATGACAGAAATCAATGGGAATCTGGGAGCCCAACACAACTCTCTGA 1655
 Db 422 AGATTATTTAAAGAAAAGATGCTGATTTGTATCAATTTTACATGGCTTGGTCTA 481
 |||||||
 QY 1656 AGATTATTTAAAGAAAAGATGCTGATTTGTATCAATTTTACATGGCTTGGTCTA 1715
 Db 482 GAGGAGGAGGCT 494
 |||||||
 QY 1716 GAGGAGGAGGCT 1728

RESULT 6
 LOCUS N28887 501 bp mRNA EST 04-JAN-1996
 DEFINITION yx63a11.r1 Homo sapiens cDNA clone 26396 5' similar to gb.M24470
 GMP REDUCTASE (HUMAN);.
 ACCESSION N28887
 NID g1147123
 KEYWORDS EST.
 SOURCE human clone-26396 primer-T7 library-Soures melanocyte 2NBHx
 vector-pT73D (Pharmacia) with a modified polylinker host-DH10B
 (ampicillin resistant) Rsite1=Not I Rsite2=Eco RI Male. 1st strand
 cDNA was primed with a Not I - oligo(dT) primer
 [5'-TGTTCAATCTGATGAGGAGGCGCGCTGTGTACTCTGAGAGAAATCTGGAGTG 917
 double-stranded cDNA was size selected, ligated to Eco RI adapters
 (Pharmacia), digested with Not I and cloned into the Not I and Eco
 RI sites of a modified pT73 vector (Pharmacia). Library
 constructed by Bento Soares and M.Fatima Bonaldo. RNA from normal
 foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P.
 Albino.

ORGANISM Homo sapiens
 Eukaryote; Metazoa; Eumetazoa; Bilateria; Coelomata;
 Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
 Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;
 Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 501)
 REFERENCE Hallier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
 AUTHORS Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
 Parsons,D., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
 Trevarakis,E., Waterston,R., Williamson,A., Wohlmann,P. and
 Wilson,R.
 TITLE The Washu-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT

Contact: Wilson RK
 Washu-Merck EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 High quality sequence stops: 372
 Source: IMAGE Consortium, LNLN
 This clone is available royalty-free through LNLN; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Location/Qualifiers
 1..501
 /organism="Homo sapiens"
 /clone="26396"
 <1..>501

BASE COUNT 120 a 116 c 130 g 134 t 1 others
 ORIGIN
 Query Match 26.5%; Score 489; DB 19; Length 501;
 Best Local Similarity 99.4%; Pred. No. 0.00e+00;
 Matches 497; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Db 1 TCCTTTTCATTTTCGGAAGCTCAAAAGCAGACATCTCTGGGTTCCCATCATTTGCGCAAT 60
 |||||||
 QY 438 TCCTTTTCATTTTCGGAAGCTCAAAAGCAGACATCTCTGGGTTCCCATCATTTGCGCAAT 497
 Db 61 ATGATATCTGTGGGCGACCTTTGAGATGGCCAAAGGTTCTGTAGTTCTCTCTCTACT 120
 |||||||
 QY 498 ATGATATCTGTGGGCGACCTTTGAGATGGCCAAAGGTTCTGTAGTTCTCTCTCTACT 557
 Db 121 GATGTCCATAAGCACTATAGCCTGTTCACTGTCGCAAGATTGCTGGCCAGAAATCTGAT 180
 |||||||
 QY 558 GCTGTCCATAAGCACTATAGCCTGTTCACTGTCGCAAGATTGCTGGCCAGAAATCTGAT 617
 Db 181 TGTCTTGAAGCATCTGGCTGCGACGCTCAGGACAGAGGCTTCTGACTTTGAGCAGCTGAA 240
 |||||||
 QY 618 TGTCTTGAAGCATCTGGCTGCGACGCTCAGGACAGAGGCTTCTGACTTTGAGCAGCTGAA 677
 Db 241 CAGATCTCGGAAGCTATCTCCAGGCTGAGATATATATGCTGATGGGCAAAATGGCTAC 300
 |||||||
 QY 678 CAGATCTCGGAAGCTATCTCCAGGCTGAGATATATATGCTGATGGGCAAAATGGCTAC 737
 Db 301 TCTGAACACTTTGTGTAATTTGTAAGATGTACGAGAGGCTTCCCGACACACCATC 360
 |||||||
 QY 738 TCTGAACACTTTGTGTAATTTGTAAGATGTACGAGAGGCTTCCCGACACACCATC 797
 Db 361 ATGCGAGGGAATGTGTACAGAGAGATGTAGAAAGCTCATCTTTCTGGGGCTGAC 420
 |||||||
 QY 798 ATGCGAGGGAATGTGTACAGAGAGATGTAGAAAGCTCATCTTTCTGGGGCTGAC 857
 Db 421 ATCATCAAGGGGAATTTGGGCGAGGCTGTGTACTACTCGGAGAAATCTGGAGTG 480
 |||||||
 QY 858 ATCATCAAGGGGAATTTGGGCGAGGCTGTGTACTACTCGGAGAAATCTGGAGTG 917
 Db 481 GGGTATCC-CAGCTCAGCGC 499
 |||||||
 QY 918 GGGTATCCACAGCTCAGCGC 937

RESULT 7
 LOCUS N30305 546 bp mRNA EST 05-JAN-1996
 DEFINITION yw69n06.s1 Homo sapiens cDNA clone 257531 3'.
 ACCESSION N30305
 NID g1148825
 KEYWORDS EST.
 SOURCE human clone-257531 primer-m13 -40 forward library-Soures placenta
 8c09weeks 2NBH8c09w vector-pT73D (Pharmacia) with a modified
 polylinker host-DH10B (ampicillin resistant) Rsite1=Not I
 Rsite2=Eco RI two placentae: one from 8 weeks and another from 9
 weeks post conception. 1st strand cDNA was primed with a Not I -
 oligo(dT) primer
 [5'-TGTTCAATCTGAAAGTGGAGGCGCGCGATTTTCTTTTCTTTT-3'],

double-stranded cDNA was size selected, ligated into Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pTZ19 vector (Pharmacia). Library constructed by Bento Soares and M. Fatima Bonaldo.

ORGANISM

Homo sapiens
Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo. 1 (bases 1 to 546)

REFERENCE

Hillier, L., Clark, N., Dubouque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

The Washu-Merck EST Project
Unpublished (1995)

TITLE

JOURNAL

Unpublished (1995)

COMMENT

Contact: Wilson RK

Washu-Merck EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

High quality sequence stops: 348

Source: IMAGE Consortium, LNL

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

Location/Qualifiers

1..546

/organism="Homo sapiens"

/clone="257531"

<1..>546

BASE COUNT 138 a 115 c 133 g 158 t 2 others

ORIGIN

Query Match 26.3%; Score 485; DB 19; Length 546;

Best Local Similarity 97.1%; Pred. No. 0.00e+00;

Matches 529; Conservative 0; Mismatches 10; Indels 6; Gaps 6;

Location/Qualifiers

1..472

/organism="Homo sapiens"

/clone="257531"

<1..>546

BASE COUNT 138 a 115 c 133 g 158 t 2 others

ORIGIN

2 ACAGCAATGAGCTCTCTTATGAGGTATTTCAATATACAAACACTTCAGAGCTTT 61

1843 ACAGCAATGAGCTCTCTTATGAGGTATTTCAATATACAAACACTTCAGAGCTTT 1784

62 TTAGATATGAGGTATTTAGTGAAGCGGATTAACAAACATGATTCYAAAAGCTT 121

1783 TTAGATATGAGGTATTTAGTGAAGCGGATTAACAAACATGATTCYAAAAGCTT 1724

122 GCCTCTCTAGACCAAGCCATGTAAGATTTATGATCAATCAGCATCTTTCCTTTT 181

1723 GCCTCTCTAGACCAAGCCATGTAAGATTTATGATCAATCAGCATCTTTCCTTTT 1664

182 AATAATCTTCAGAGTGTGTGGTGGTCCCGAGATTCATTTGCTTCATCATT 241

1663 AATAATCTTCAGAGTGTGTGGTGGTCCCGAGATTCATTTGCTTCATCATT 1604

242 TTCCTCAGACTGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 301

1603 TTCCTCAGACTGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1544

302 GCAGTATAGAGTACAGAGTACTGACAGAGCCCTCAGAGATGAGACAAAGAAATGA 361

1543 GCAGTATAGAGTACAGAGTACTGACAGAGCCCTCAGAGATGAGACAAAGAAATGA 1484

362 CAAATATATACAGAGTACAGAGTGGATGGATGGATGGATGGATGGATGGATGGAT 420

1483 CAAATATATACAGAGTACAGAGTGGATGGATGGATGGATGGATGGATGGATGGAT 1424

421 GGTAGAGTACTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 480

1423 GGTAGAGTACTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1367

Db 481 GAGANTGATTCACCTGCTGGTGAATCCCGAGTGAAGTATGTTCTCCNGGCTAATCCT 540

Cp 1366 AAGATGGATTCACCTGCTGGTGAATCCCGAGTGAAGTATGTTCTCCNGGCTAATCCT 1309

Db 541 TGAGC 545

Cp 1308 TGAGC 1304

RESULT 8

LOCUS AA778158 472 bp mRNA EST 05-FEB-1998

DEFINITION z445b09.s1 Soares fetal heart NBH19W Homo sapiens cDNA clone

379865.3 similar to gb:M24470 GMP REDUCTASE (HUMAN);, mRNA

sequence.

ACCESSION AA778158

NID 92837559

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 472)

Hillier, L., Allen, M., Bowles, L., Dubouque, T., Gelsel, G., Jost, S.,

Kitzman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,

Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,

Theising, B., White, F., Wylie, T., Waterston, R. and Wilson, R.

Washu-Merck EST Project

Unpublished (1997)

CONTACT: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

Seq primer: -40m13 fwd. ET from Amersham

High quality sequence stop: 456.

Location/Qualifiers

1..472

/organism="Homo sapiens"

/note="Organ: heart; Vector: pTZ193D (Pharmacia) with a

modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer (5'

TGTTCACCAATCTGAGAGCGGCGCCGATCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT

3'), double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified pTZ193 vector

(Pharmacia). Library went through one round of

normalization to a Cot = 5. Library constructed by

M. Fatima Bonaldo. This library was constructed from the

same fetus as the fetal lung library, Soares fetal lung

NBH19W."

/db_xref="GDB:1288121"

/db_xref="taxon:9606"

/clone="379865"

/clone_1bp="Soares fetal heart NBH19W"

/sex="unknown"

/dev_stage="19 weeks"

/lab_host="DH10B (ampicillin resistant)"

BASE COUNT 101 a 113 c 124 g 134 t

ORIGIN

Query Match 25.2%; Score 464; DB 13; Length 472;

Best Local Similarity 99.2%; Pred. No. 0.00e+00;

Matches 468; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Location/Qualifiers

1..472

/organism="Homo sapiens"

/note="Organ: heart; Vector: pTZ193D (Pharmacia) with a

modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer (5'

TGTTCACCAATCTGAGAGCGGCGCCGATCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT

3'), double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified pTZ193 vector

(Pharmacia). Library went through one round of

normalization to a Cot = 5. Library constructed by

M. Fatima Bonaldo. This library was constructed from the

same fetus as the fetal lung library, Soares fetal lung

NBH19W."

Cp	1595	CTGTTTGGTCCTCCTCCACAGTAGGCCCTTGCGCATTTTGTGTGCAGATAG	1536
Dp	121	AGAACTACAGATTACTGCAGAGGCCCTCAGAGTAGAACAACAGGAATGCAAGAATA	180
Cp	1555	AGAACTACAGATTACTGCAGAGGCCCTCAGAGTAGAACAACAGGAATGCAAGAATA	1476
Dp	181	TACAGAGCTGCAGTAGCTGGGATGGGTATAGAACCCCACTTGGGATGCCCATGGTAGT	240
Cp	1475	TACAGAGCTGCAGTAGCTGGGATGGGTATAGAACCCCACTTGGGATGCCCATGGTAGT	1416
Dp	241	ACTGGTGCCTTGGAGGGGTAGAACTGCTCAGGTCTAGCACGCCCTCCTCACTGAAGATTGAT	300
Cp	1415	ACTGGTGCCTTGGAGGGGTAGAACTGCTCAGGTCTAGCACGCCCTCCTCACTGAAGATTGAT	1356
Cp	1355	CACCCTGCGGTGATCTCGAATGAGGTGTTCTCTGCTCAACTTTGAGCTTAGCTGC	1296
Dp	361	TCCCACATAGGTAACGCTAGAGCGGATCCCTCCTAGATGCTCGAGTATGTTCCAC	420
Cp	1235	TCCCACATAGGTAACGCTAGAGCGGATCCCTCCTAGATGCTCGAGTATGTTCCAC	1236
Dp	421	ATTCCTCTTAAAGAACTTCCACATGCTTTTCCCTCCTAGAGCTCTGATCTCA	472
Cp	1235	ATTCCTCTTAAAGAACTTCCACATGCTTTTCCCTCCTAGAGCTCTGATCTCA	1184
RESULT	9		
LOCUS	N35234	474 bp	mRNA EST 16-JAN-1996
DEFINITION	xy21d05.s1 Homo sapiens cDNA clone 271881 3'.		
ACCESSION	N35234		
KEYWORDS	g1156376		
SOURCE			
REFERENCE			
AUTHORS			
ORGANISM	Homo sapiens		
COMMENT	Eukaryotes; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Saccophrygillia; Choanata; Tetrastoda; Amphibia; Mammalia; Theria; Euthera; Archonta; Primates; Catarrhini; Hominoidea; Homo. 1 (bases 1 to 474) Hallier,L., Clark,N., Dubube,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le.M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R. The Mashu-Merck EST Project Unpublished (1995)		
TITLE	Contact: Wilson RK		
JOURNAL	Mashu-Merck EST Project		
COMMENT	Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu High quality sequence stops: 384 Source: IMAGE Consortium, LNLN This clone is available royalty-free through LNLN ; contact the IMAGE Consortium (info@image.lnl.gov) for further information. Location/qualifiers . . .474 /organism="Homo sapiens"		
FEATURES			
Source			

BASE COUNT	118 a	101 c	117 g	137 t	1 others
MRNA	/clone="271881"				
ORIGIN	<1. ->474				
Query Match	25.2%; Score 465; DB 19; Length 474;				
Best Local Similarity	99.4%; Pred. NO. 0.00e+00;				
Matches 472; Conservative	0; Mismatches 2; Indels 1; Gaps 1;				
Db	1	TTAGATATGTGAAGGTCCAAATTAGTGAAGCGGATTAACAACAATGATGTTCTAAAGCCT	60		
Cp	1783	TTAGATATGTGAAGGTCCAAATTAGTGAAGCGGATTAACAACAATGATGTTCTAAAGCCT	1724		
Db	61	GCCCTCTTACACCAAGGCCATGTAAGATTTATGATGCAATGACATCTTTCCCTTT	120		
Cp	1723	GCCCTCTTACACCAAGGCCATGTAAGATTTATGATGCAATGACATCTTTCCCTTT	1664		
Db	121	AATAATCTTCAGAGATGTTGTGTTGGGTCGCCAGATATCCATTTGATTTCTTGACATCT	180		
Cp	1663	AATAATCTTCAGAGATGTTGTGTTGGGTCGCCAGATATCCATTTGATTTCTTGACATCT	1604		
Db	181	TTCTCAGACTGTGTCCTTCTCTTCTCTCCCAAGTACGTCCTTGGN-AATTTGTGTGT	239		
Cp	1603	TTCTCAGACTGTGTCCTTCTCTTCTCTCCCAAGTACGTCCTTGGGCAATTTTGTGTGT	1544		
Db	240	GCAGATATAGAAATGATACAGATTTACTGCAGAGAGCCCTCAGAGATGAGACAAAGAAAGA	299		
Cp	1543	GCAGATATAGAAATGATACAGATTTACTGCAGAGAGCCCTCAGAGATGAGACAAAGAAAGA	1484		
Db	340	CAAAATATATACAGAGACTCAGTACGTGGGATGGGATGAGAGACCCCACTGGGATGCCCAT	359		
Cp	1483	CAAAATATATACAGAGACTCAGTACGTGGGATGGGATGAGAGACCCCACTGGGATGCCCAT	1424		
Db	360	GGTAGAGTACTGTTGCTTGGGAGGATGAACTGCTCAGTCTAGACAGCCCTCACTGAAG	419		
Cp	1423	GGTAGAGTACTGTTGCTTGGGAGGATGAACTGCTCAGTCTAGACAGCCCTCACTGAAG	1364		
Db	420	AATGATTCACCTGCTGCTGGGTGACCTGCGATGAAGTAGTTCTTCCTGCTCACTCTT	474		
Cp	1363	AATGATTCACCTGCTGCTGGGTGACCTGCGATGAAGTAGTTCTTCCTGCTCACTCTT	1309		
RESULT	10				
LOCUS	AA456771	478 bp	MRNA	EST	06-JUN-1997
DEFINITION	z=27f02.21 Soares ovary tumor NDHOT Homo sapiens cDNA clone 770523				
ACCESSION	AA456771				
NID	52177192				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
AUTHORS	1 (bases 1 to 478)				
	Hillier,L., Allen,M., Bowles,L., Dubuque,T., Gelsel,G., Jost,S., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Scheilenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.				
TITLE	Washu-Merck EST Project 1997				
JOURNAL	Unpublished (1997)				
COMMENT	Contact: Wilson RK Washu-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu This clone is available royalty-free through LNC ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: -28n13 rev2 ET from Amersham. Location/Oualifiers				

```

source
1. 478
/organism="Homo sapiens"
/note="Organ: ovary; Vector: p1773D (Pharmacia) with a
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dt) primer (5
TGTTCACATCTGAAGTGGAGCGGGCGGGTTTTTTTTTTTTTTTTTTT 3')/
TGTTCACATCTGAAGTGGAGCGGGCGGGTTTTTTTTTTTTTTTTTTT 3')/
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p1773 vector
(Pharmacia). Library constructed by Bento Soares and
M.Fatima Bonaldo."
/db_xref="taxon:9606"
/clone="770523"
/clone_11b="Soares ovary tumor NBH07"
/sex="Female"
/tissue_type="Ovarian tumor"
/lab_host="DH10B (ampicillin resistant)"
<1. >478
<1. >478
/db_xref="GDB:5980396"
BASE COUNT 131 a 122 c 122 g 103 t
ORIGIN
Query Match 25.2% Score 465 DB 7: Length 478:
Best Local Similarity 99.8% Pred. No. 0.00e+00:
Matches 466; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 12 TGGGGCGTGCTGAGTACAGAGCCTCAGAGGGAAGACAGTGGAGTCTTTTAAAG 71
Qy 1172 TGGGGCGTGCTGAGTACAGAGCCTCAGAGGGAAGACAGTGGAGTCTTTTAAAG 1231

Db 72 AAGATGGGAACATACCATCCGAGACATCCTTAGAGAGGATCCGCTTACGTACTATGT 131
Qy 1232 AAGATGGGAACATACCATCCGAGACATCCTTAGAGAGGATCCGCTTACGTACTATGT 1291

Db 132 GGGAGCAGCTAAGCTCAAGAGTTGAGCAGAGAACTACCTTATCCGAGTACCCAGCA 191
Qy 1292 GGGAGCAGCTAAGCTCAAGAGTTGAGCAGAGAACTACCTTATCCGAGTACCCAGCA 1351

Db 192 GGTGAATCCAACTTCAGTAGAGCGTGTAGACCTGAGCAAGTCTTACCTCCAAAGCAC 251
Qy 1352 GGTGAATCCAACTTCAGTAGAGCGTGTAGACCTGAGCAAGTCTTACCTCCAAAGCAC 1411

Db 252 CAGTACTTACCATTTGGGGATCCCAAGTGGGTCCTTACCCATCCAGTACTGACGCTC 311
Qy 1412 CAGTACTTACCATTTGGGGATCCCAAGTGGGTCCTTACCCATCCAGTACTGACGCTC 1471

Db 312 TGTATACTTTGTCATTTCTGTCATCTGTCATCCTGTGAGGCTCCTGCAGTACTGTAC 371
Qy 1472 TGTATACTTTGTCATTTCTGTCATCTGTCATCCTGTGAGGCTCCTGCAGTACTGTAC 1531

Db 372 TTCTCATTTGCAACACAAATGCCCCAAGGCACTCACTGGGAGGAAGAAGAACAA 431
Qy 1532 TTCTCATTTGCAACACAAATGCCCCAAGGCACTCACTGGGAGGAAGAAGAACAA 1591

Db 432 ACAGTGTGAGAAAATGATGCAGAAAATCAATATGGCAATCTGGGGAC 478
Qy 1592 ACAGTGTGAGAAAATGATGCAGAAAATCAATATGGCAATCTGGGGAC 1638

RESULT 11
LOCUS AA574317 599 bp mRNA EST 21-AUG-1997
DEFINITION ng33h05.s1 NCI-CGAP_C03 Homo sapiens cDNA clone IMAGE:936633, MRNA
sequence.
ACCESSION AA574317
NID g2265245
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Eukaryota; Eukaryota; Eukaryota; Metazoa; Chordata;
Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea;
Homo.
1 (bases 1 to 599)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

```

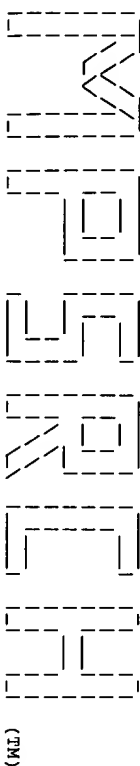
TYPE	Journal	COMMENT
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
Unpublished (1997)		
Contact: Robert Strausberg, Ph.D.		
Tel: (301) 496-1550		
Email: Robert_Strausberg@nih.gov		
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D., Ph.D.		
CDNA Library Preparation: M. Bento Soares, Ph.D.		
CDNA Library Arraying: Greg Lennon, Ph.D.		
DNA Sequencing by: Washington University Genome Sequencing Center		
Clone distribution: NCI-CGAP clone distribution Information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bdrrp/image/image.html		
Insert length: 590	Std Error: 0.00	
Seq primer: -40ml3 fwd. RT from Amersham		
High quality sequence stop: 275.		
Location/Qualifiers		
1..599		
/organism="Homo sapiens"		
/note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from 12 pooled bulk tumor samples and primed with a Not I - Oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT3 vector. Library went through one round of normalization."		
/db_xref="taxon:9606"		
/clone="IMAGE:936633"		
/clone_1id="NCI_CGAP_Co3"		
/sex="pooled"		
/tissue_type="colon"		
/lab_host="DH10B"		
<1..>599		
BASE COUNT	156 a 129 c 142 g 171 t 1 others	
ORIGIN		
mrna		
Query Match	25.0%; Score 460; DB 9; Length 599;	
Best Local Similarity 97.7%;	Pred. No. 0.00e+00;	
Matches 501; Conservative	0; Mismatches 6; Indels 6; Gaps 6;	
1 ACAGCAATGAGAGCTCTCTTTATGAGGATTTCAATATACAAACACTCAGAGCTTT		60
CP 1843 ACAGCAATGAGAGCTCTCTTTATGAGGATTTCAATATACAAACACTCAGAGCTTT		1784
DB 61 TTGATATGTGAAGGTCCAAATTTAGTGAAGGGATTAACAACATGATTTCAAAAGCCT		120
CP 1783 TTGATATGTGAAGGTCCAAATTTAGTGAAGGGATTAACAACATGATTTCAAAAGCCT		1724
DB 121 GCCCTCCTAGACCAAGGCATGTAAAGATTTATGTACCAATCAGCATCTTTCTTTT		180
CP 1723 GCCCTCCTAGACCAAGGCATGTAAAGATTTATGTACCAATCAGCATCTTTCTTTT		1664
DB 181 AATAATCTTCAGAGATGTGTGTGGGTCCCAAGATTCACATTTGATTTCTTGATCATTT		240
CP 1663 AATAATCTTCAGAGATGTGTGTGGGTCCCAAGATTCACATTTGATTTCTTGATCATTT		1604
DB 241 TTCGAGAGCTGTTTGCTTCTTGCTTCTCCCAAGTGAAGTGCCTTGAGGATTTGTGTGT		300
CP 1603 TTCGAGAGCTGTTTGCTTCTTGCTTCTCCCAAGTGAAGTGCCTTGAGGATTTGTGTGT		1544
DB 301 GCAGATAGAGAAATACAGATTACTGACGAGAGCCCTCAGAGTGAACACAGGAATGA		360
CP 1543 GCAGATAGAGAAATACAGATTACTGACGAGAGCCCTCAGAGTGAACACAGGAATGA		1484
DB 361 NCAATTAATACAGAGCTGACAGTGGGATGGGAGAGAGCCCACTGGGATGCCCAT		420
CP 1483 CAAATTAATACAGAGCTGACAGTGGGATGGGAGAGAGCCCACTGGGATGCCCAT		1424
DB 421 GGTAAAGTACTGTGC-TTGGAGAGGA-AACTGCTCAGGTC-AGCAGGCC-CACTGGAG		476

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 474)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.11nl.gov/dbirp/image/image.html
Insert Length: 1720 Std Error: 0.00
Seq primer: -40ml3 fwd. RT from Amersham.
Location/Qualifiers
1. 474
/organism="Homo sapiens"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCGAAGTGGGAGCGGCCGCTCATTTTCTTTTCTTTT-
3']. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
/db_xref="taxon:9606"
/clone="IMAGE:1353701"
/clone_lib="NCI-CGAP-GCBI"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
BASE COUNT 125 a 93 c 109 g 147 t
ORIGIN
Query Match 24.1%; Score 445; DB 16; Length 474;
Best Local Similarity 98.7%; Pred. No. 0.00e+00;
Matches 461; Conservative 0; Mismatches 4; Indels 2; Gaps 2;
Db 10 ACAGTCAATAGAGCTCTCTCTTATGTAGGTATTTCAATATACAAACACTTCAGAGCTTT 69
CP 1843 ACAGTCAATAGAGCTCTCTCTTATGTAGGTATTTCAATATACAAACACTTCAGAGCTTT 1784
Db 70 TTAGATATGTGAAGGTCCATTTAGTGAAGCGATTAAACAACATGATCTTAAAGCCT 129
CP 1783 TTAGATATGTGAAGGTCCATTTAGTGAAGCGATTAAACAACATGATCTTAAAGCCT 1724
Db 130 GCCTCCTAGACCAAGCCATGTAAAGATTTATGTACCAATCAGCATCTTTCCTTTT 189
CP 1723 GCCTCCTAGACCAAGCCATGTAAAGATTTATGTACCAATCAGCATCTTTCCTTTT 1664
Db 190 AATAATCTTAGAGATGTGTGTGGTCCCAATTTCCATTTGATTTCTTGATCATTT 249
CP 1663 AATAATCTTAGAGATGTGTGTGGTCCCAATTTCCATTTGATTTCTTGATCATTT 1604
Db 250 TTCTCAGACTGTTGCTTCTTCTTCTCCCAAGTGAAGTGGATATTGTTGTGT 309
CP 1603 TTCTCAGACTGTTGCTTCTTCTTCTCCCAAGTGAAGTGGATATTGTTGTGT 1544
Db 310 GCAGATAGAGAGTACAGAGTACTGCAGAGGCCCTCAGAGTGAACAACAGGAATGA 369

CP 1543 GCAGATAGAGAGTACAGAGTACTGCAGAGGCCCTCAGAGTGAACAACAGGAATGA 1484
Db 370 CAAGTAAATACAGAGTGCAGTACGTGGATGGGTAGAGACCC-ACCTGGATGCC-AT 427
CP 1483 CAAGTAAATACAGAGTGCAGTACGTGGATGGGTAGAGACCCACCTGGATGCCCAT 1424
Db 428 GGTAGAGTACTGCTGCTTGGAGGAGTGAACCTCTCAGGCTAGCA 474
CP 1423 GGTAGAGTACTGCTGCTTGGAGGAGTGAACCTCTCAGGCTAGCA 1377

Search completed: Wed Jun 17 13:00:01 1998
Job time : 3135 secs.

This Page Blank (uspto)



Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MParch_dp protein - protein database search, using Smith-Waterman algorithm
Run on: Mon Jun 15 15:56:34 1998; Maspar time 10.68 Seconds
525.802 Million cell updates/sec
Tabular output not generated.

Title: >US-08-774-169-1
Description: (1-366) from US08774169.pep
Perfect Score: 2632
Sequence: 1 MTCCLPALRFIATPRLSAMP.....RRTFIRVYQGVNPIFSEAC 366

Scoring table: PAM 150
Gap 11

Searched: 124785 seqs, 15338987 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseg31-2
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27

Statistics: Mean 34.191; Variance 147.545; scale 0.232

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	349	13.3	514	1	Human IMPDH.	3.84e-21
2	347	13.2	514	1	Chinese hamster IMPDH	5.87e-21
3	112	4.3	438	25	Staphylococcus aureus	1.28e+00
4	111	4.2	487	23	Sugar biosynthesis en	1.52e+00
5	101	3.8	168	27	E. coli colonisation	8.30e+00
6	97	3.7	661	9	Enzyme involved in ei	1.61e+01
7	97	3.7	661	20	Biosynthetic enzyme o	4.26e+01
8	91	3.5	256	19	Helicobacter pylori 2	3.63e+01
9	92	3.5	265	2	Mouse serine protease	4.26e+01
10	91	3.5	267	2	Sequence of serine pr	4.26e+01
11	92	3.5	422	7	ZmRL1 homologue prote	3.63e+01
12	92	3.5	520	25	Arabidopsis thaliana	5.00e+01
13	90	3.4	216	26	C. glutamicum orf3 pr	5.86e+01
14	89	3.4	307	16	Novel human phosphoy	5.86e+01
15	89	3.4	359	20	Prostaglandin DP rece	5.86e+01
16	89	3.4	558	7	XRS	5.86e+01
17	89	3.4	2408	2	Translation of ORF 3	5.86e+01
18	88	3.3	202	2	Glycopeptide resistan	6.87e+01
19	87	3.3	289	27	Rat syntaxin 3 protei	8.04e+01

20	87	3.3	343	5	R28887	PR18 fusion protein.	8.04e+01
21	88	3.3	403	13	R71933	Cobra venom factor C-	6.87e+01
22	87	3.3	530	5	R28888	R1 paraneoplastic ant	8.04e+01
23	86	3.3	750	1	P82940	Artrial natriuretic po	9.40e+01
24	85	3.2	211	18	W00581	Programmed mammalian	1.10e+02
25	85	3.2	211	12	R60867	RP-2 programmed cell	1.10e+02
26	84	3.2	255	6	R30808	3-alpha-hydroxy-ster	1.28e+02
27	83	3.2	286	17	R88917	Modified incp protein	1.50e+02
28	84	3.2	305	25	W23077	Sulfolobus solfataric	1.28e+02
29	85	3.2	376	15	R80016	Marmoset alpha-1,3-ga	1.10e+02
30	85	3.2	399	19	W04215	Rat vas deferens p2x	1.10e+02
31	85	3.2	399	19	W04218	Human urinary bladder	1.10e+02
32	83	3.2	421	14	R71416	C. glutamicum protein	1.50e+02
33	83	3.2	421	12	R63566	Feedback inhibition r	1.50e+02
34	83	3.2	421	23	W06586	B. lactofermentum asp	1.50e+02
35	83	3.2	421	15	R86307	Corynebacterium gluta	1.50e+02
36	84	3.2	491	18	R82243	Human fibrinogen B-be	1.28e+02
37	85	3.2	562	12	R63588	Full length HA protei	1.10e+02
38	84	3.2	572	8	R40843	Bilirubin oxidase.	1.28e+02
39	85	3.2	748	10	R52706	Human enkephalinase.	1.10e+02
40	85	3.2	750	10	R52707	Rat enkephalinase.	1.10e+02
41	85	3.2	750	1	P82867	Enkephalinase (rat).	1.10e+02
42	83	3.2	842	12	R63570	Feedback inhibition r	1.50e+02
43	83	3.2	842	12	R63567	Feedback inhibition r	1.50e+02
44	84	3.2	950	2	R06376	Product of the sscI g	1.28e+02
45	85	3.2	1266	24	W25157	Tomato immunity 2 (fu	1.10e+02

ALIGNMENTS

RESULT ID	1	ALIGNMENTS
AC	R05432; standard; protein; 514 AA.	
AC	R05432; 31-AUG-1990 (first entry)	
DE	Human IMPDH.	
KW	Inosine 5'-mono-phosphate dehydrogenase; hepatomas;	
OS	Homo sapiens.	
PN	MO9001545-A.	
PD	22-FEB-1990.	
PF	02-AUG-1989; U0344.	
PR	12-AUG-1988; US-232302.	
PA	(ARCH-) Arch. Dev. Corp.	
PI	Collart FR, Huberman E;	
DR	WPI: 90-083504/11.	
DR	P-RSDB; R05432.	
PT	DNA encoding eukaryotic inosine 5'-mono-phosphate dehydrogenase -	
PT	used to detect high levels of enzyme activity of tumour cells,	
PS	esp. hepatomas, and to produce guanosine monophosphate.	
CC	A 35 aa segment corresponds to deduced aa residues 336-370 in	
CC	both the human and Chinese hamster proteins.	
CC	See also Q03541.	
SQ	Sequence 514 AA:	

Query Match 13.3%; Score 349; DB 1; Length 514;
Best Local Similarity 30.3%; Pred. No. 3.84e-21;
Matches 71; Conservative 66; Mismatches 93; Indels 4; Gaps 4;

Db	231	rdyplaaekdkkqlcgaaigtbeddkyrlidlagaqvdvvlssqgnslfimmkyl 290	
Qy	104	OEFGAGNPDCLEHLAASSGSSDFEQLEILAIPOVKYICIDVANGVSEHVEFVKDV 163	
Db	291	kdyrplqivlgvgnvtaagknildagvdalrygmssgscitqevlaagrpqatavkyv 350	
Qy	164	RRRPPQHTIMAGNVYGEWEELILSGADILIKVIGISVCTRRKRTGVGVPDLSAVMEC 223	
Db	351	seyarfygvviadggqigvghakalalagstwmmsllaateagpgevfedqyrlkk 410	
Qy	224	ADAAGKIGKHIISDGGSCGVDYAKAFAGADRVMLGMLAGSSESGELIENDGKKYL 283	
Db	411	yrmgslamdckhlssqnyr-fseadkikva-gg-vsgavqdk-gslhkfvpyl 460	

OY 284 FYGMSSEAMKRYAGVAEYRASSEGKTEVPFKGDVEHTIRDLIGIRSTCTYV 337

RESULT 2

ID R05431 standard; protein; 514 AA.

AC R05431;

DT 31-AUG-1990 (first entry)

DE Chinese hamster IMPDH.

KW Inosine 5'-mono-phosphate dehydrogenase; hepatomas;

KM guanosine monophosphate.

OS Cricetus sp.

PN W09001545-A.

PD 22-FEB-1990.

PE 02-AUG-1989; U0344.

PR 12-AUG-1988; US-232302.

PA (ARCH-) Arch. Dev. Corp.

PI Collart FR, Huberman E;

DR N-PSDB: 003541.

PT DNA encoding eukaryotic inosine 5'-mono-phosphate dehydrogenase -

PT used to detect high levels of enzyme activity of tumour cells,

PT esp. hepatomas, and to produce guanosine monophosphate.

PS Claim 8; Fig 2; 51pp; English.

CC A 35 aa segment corresponds to deduced aa residues 336-370 in

CC both the human and Chinese hamster proteins.

CC See also Q03540.

SO Sequence 514 AA;

Query Match 13.2%; Score 347; DB 1; Length 514;

Best Local Similarity 30.3%; Pred. No. 5,876-21;

Matches 71; Conservative 66; Mismatches 93; Indels 4; Gaps 4;

Db 231 rdyplaskdakkqlcgaaigttheddkyrlallagaydvvnvdsggnsifqimlkym 290

OY 104 QEFAGQNPDCLEHLAASSGTSSEDFEQLLEALPQVKYICLDVANGYSEHFEFVNDV 163

Db 291 kekynlqvlgnvntvtaqaknldagvdalrvmgcsicltgcvlaagcpataaykv 350

OY 164 RKRPPQHTIMAGNVVTGMEBELISGADITKVGIGPSSVCTTRKKTGVGPQSAVMEC 223

Db 351 seayrtfvpvrladgqlqnvghlakalagastvmmgsllaateapeyffsdqrlrk 410

OY 224 ADAHGLKGIHTISDGGSCPGDVAKAFGAGADFYMLGMLAGHSGGELLERDCKKTL 283

Db 411 yrgngslamdnhlsqnrty-fseadkikva-qg-vsgavgdk-gslkftpyl 460

OY 284 FYGMSSEAMKRYAGVAEYRASSEGKTEVPFKGDVEHTIRDLIGIRSTCTYV 337

RESULT 3

ID W18209 standard; protein; 438 AA.

AC W18209;

DT 19-JAN-1998 (first entry)

DE Staphylococcus aureus Coenzyme A disulphide reductase (COADR).

KW S. aureus; COADR; Coenzyme A disulphide reductase; diagnosis; pneumonia;

KW antimicrobial; antibodies; staphylococcal; enterococcal; meningitis;

KW gastroenteritis; dermal infection; Gram positive bacterial infection;

KM cerebritis; food poisoning.

OS Staphylococcus aureus.

PH Key Location/Qualifiers

FT Misc-difference 178

PN W09723628-A1.

PD 03-JUL-1997.

PE 19-DEC-1996; U20017.

PR 22-DEC-1995; US-009146.

PA (UYBR-) UNIV BRITISH COLUMBIA.

PI Davies JE, Delcardayre SB;

DR WPI: 97-351064/32.

DR N-PSDB: T71106.

PT New isolated Staphylococcus aureus Coenzyme A disulphide reductase

PT - used to develop products for treating Gram-positive bacteria

PT infections and for detection and diagnosis

PS Claim 2; Fig 2; 48pp; English.

CC This is an isolated Staphylococcus aureus Coenzyme A disulphide reductase

CC (COADR). COADR catalyses the specific reduction of Coenzyme A disulphide

CC to Coenzyme A with the concomitant oxidation of NADPH to NADP. The DNA

CC encoding COADR can be obtained from an S. aureus genomic DNA library

CC using PCR primers with the standard techniques. COADR was fractionated

CC from cellular extracts of S. aureus by following the NADPH- and CoA

CC disulphide dependant reduction of dihydro-bis-2-nitrobenzoic acid (DTNB).

CC The S. aureus COADR can be used to identify compounds which can inhibit

CC the enzyme. Such compounds can be used to treat Gram-positive bacterial

CC infections such as staphylococcal and enterococcal infections. They can

CC be used for treating gastroenteritis, toxic shock syndrome, scalded skin

CC syndrome, dermal infections, bone and joint infections, pneumonia and

CC emphysema, meningitis, cerebritis, endocarditis, bacteraemia, septic

CC shock, septicemia, food poisoning, or enteritis. The products can also

CC be used for detection and diagnosis.

SO Sequence 438 AA;

Query Match 4.3%; Score 112; DB 25; Length 438;

Best Local Similarity 27.9%; Pred. No. 1,286+00;

Matches 34; Conservative 26; Mismatches 51; Indels 11; Gaps 8;

Db 145 anygdkylvvgagysvle-vlenenrglhpcllnrdklnkmdamngplldelk 203

OY 243 PGDVAKKFGAGADRYMLGMLAGHSGSG--GELLERDCKTKKLFYGMSSEAMKRYAGV 300

Db 204 lpyrlneelnaingnektfsgkvehymdllegvgtlpnaskflessnik-ldtkg-fipv 261

OY 301 AEYRASSEG-KTV--EVPFK-GDVEHTIRDLIG-GIRSTCTYGAAKRLKLSRTTPIRV 354

Db 262 nd 263

OY 355 TQ 356

RESULT 4

ID W19740 standard; protein; 487 AA.

AC W19740;

DT 18-SEP-1997 (first entry)

DE Sugar biosynthesis enzyme EryBVI.

KW Polyketide; glycosylation; eryBVI; L-mycarose; antimicrobial;

KW antibiotic; antifungal; fungicide; anticancer; cytostatic;

KM anthelmintic.

OS Saccharopolyspora erythraea.

PN W09723630-A2.

PD 03-JUL-1997.

PE 23-DEC-1996; U20238.

PR 21-DEC-1995; US-576626.

PA (ABBO) ABBOTT LAB;

PI Donadio S, Katz L, Staver MJ, Summers RG;

DR WPI: 97-351066/32.

DR N-PSDB: T72685.

PT New genes involved in sugar biosynthesis and attachment - used to

PT generate polyketide antimicrobials etc. with altered pattern of

PT glycosylation

PS Disclosure; Fig 4B; 85pp; English.

CC EryBVI (W19740) is an enzyme involved in the biosynthesis of the

CC sugar L-mycarose. It is one of 10 enzymes (see also W19734-39 and

CC W19741-43) predicted to be involved in D-desosamine or L-mycarose

CC biosynthesis and attachment that have been identified from gene

CC clusters (T72684 and T72685) of Saccharopolyspora erythraea. Novel

CC glycosylation-modified polyketides are produced by selectively

CC altering, inactivating or augmenting the eryB and/or eryC genes

CC encoding these sugar biosynthesis enzymes and introducing them into

CC polyketide-producing microorganisms.

SO Sequence 487 AA;

Query Match 4.2%; Score 111; DB 23; Length 487;

Best Local Similarity 28.4%; Pred. No. 1,526+00;

Matches 21; Conservative 22; Mismatches 28; Indels 3; Gaps 3;

Db 203 pntfwlvagqlmhlhndvnmldrlvrlacvptavrdtraddvlarlpgsfqarlhs 262

QY 6 PALRFLATPLSLSMPIHDVYKLDKEDVLLRPKRSTLKSREVDLTR-S-FSFRNSK-OT 62

DB 263 ffgagtppannmsl 276

QY 63 YSGVPPIIANNMTIV 76

ID RESULT 5
 AC W38342; standard; Protein; 168 AA.
 DT 27-MAR-1998 (first entry)
 DE E. coli colonisation factor antigen CS3.
 KW Bacterial colonisation; colonisation factor antigen; CS3;
 KW enterotoxigenic Escherichia coli; vaccine; diagnosis; research.
 OS Escherichia coli.
 PN US5698416-A.
 PD 16-DEC-1997.
 PF 02-JUN-1995; 460739.
 PR 02-JUN-1995; US-460739.
 PA (USA) US SEC OF ARMY.
 PI Bell BA, Cassels FJ, Wolf MK;
 DR WPI: 98-051486/05.
 DR N-PSDB; T96060.
 PT Production of bacterial colonisation factor protein - by expression
 PT under control of heat-inducible promoter
 PS Example 2; Columns 17-20; 11pp; English.
 CC Production of a protein that affects bacterial colonisation.
 CC CC comprises inoculating a broth containing tryptone and yeast extract
 CC with enteric bacteria containing a DNA sequence encoding the
 CC protein under the control of a temperature regulated promoter,
 CC culturing the bacteria, removing the bacteria from the medium and
 CC recovering the protein. The method is used especially for producing
 CC the colonisation factor antigen CS3 of enterotoxigenic E. coli, i.e.
 CC the antigen denoted by the present sequence, which may be used in
 CC vaccines or for diagnostic or research purposes. Growing the
 CC bacteria at low temperature until the late logarithmic phase
 CC increases the yield of the protein.
 QO Sequence 168 AA;

RESULT	6	standard; Protein: 661 AA.
ID	RA2450	
AC	R42450;	
DT	27-MAY-1994	(first entry)
DE	Enzyme Involved In eicosapentaenoic acid (EPA) synthesis.	
KW	EPA: eicosapentaenoic acid synthetase; drug: anticoagulant;	
KM	hypolipemic; hypolipemic; antihypertensive; anticancer; pesticide;	
KW	foodstuff; additive.	
OS	Shewanella putrefaciens.	
PN	W09323545-A.	
PD	25-NOV-1993.	
PF	14-MAY-1993: J00641.	
PR	15-MAY-1992; JP-147945.	
PA	(SAGA) SAGAMI CHEM RES CENTRE.	
PI	Kato S, Kondo K, Yamada A, Yazawa K;	
DR	WPI: 93-386577/48.	
DR	N-PSDB: 051128.	
PT	Gene coding for eicosapentaenoic acid synthetase - Is isolated	
PT	from Pseudomonas, Alteromonas or Shewanella and used for	
PT	recombinant prodn. of eicosapentaenoic acid	
PS	Claim 4; Page 40-44; 106PP; Japanese.	
CC	EPA is useful as a drug, having anticoagulant, hypolipemic,	
CC	hypolipemic, antihypertensive and anticancer activity. It is also	
CC	a pesticide and is useful as a nutritional foodstuff and animal feed	
CC	additive.	

SQ	Sequence	661 AA;
	Query Match	3.7%;
	Best Local Similarity	23.4%;
Matches	11; Conservative	19; Mismatches 16; Indels 1; Gaps 1.

RESULT 7
 ID R99458 standard; Protein: 661 AA.
 AC R99458;
 DT 30-JAN-1997 (first entry)
 DE Biosynthetic enzyme of Icosapentaenoic acid synthase.
 DE Icosapentaenoic acid synthase; EPA: drugs: agrochemicals;
 KW foodstuffs; animal feed; lipid balance correction; antihypertensive;
 KW antiinflammatory; anticancer agent.
 OS *Shewanella putrefaciens*.
 PN M09621735-A1.
 PD 18-JUL-1996.
 PF 12-JAN-1996; J00030.
 PR 13-JAN-1995; JP-004299.
 PA (SAGA) SAGAMI CHEM RES CENTRE.
 PI Kato S, Kondo K, Yamada A, Yazawa K;
 PI WPI: 96-342288/34.
 DR N-PSDB; T34137.
 PT Production of Icosapentaenoic acid using transformed *E. coli* - uses
 PT DNA coding for Icosapentaenoic acid synthase derived from *Shewanella*
 PT strain
 PS Disclosure: Page 45-47; 145pp: English.
 CC The DNA sequence (T34137) which encodes the biosynthetic enzymes of
 CC Icosapentaenoic acid (EPA) can be used to transform *Escherichia coli*
 CC The DNA sequence allows efficient microbial production of EPA, which
 CC is a raw material for drugs, agrochemicals, foods and animal
 CC feedstuffs. EPA is also useful for lipid balance correction and as
 CC an antihypertensive, antiinflammatory and anticancer agent.
 Q0 Sequence 661 AA;

RESULT 8
ID W05197 standard; Protein; 256 AA.
AC W05197;
DT 31-JAN-1997 (first entry)
DE Helicobacter pylori 29 kDa protective antigen H5.1.
KW Antigen; vaccine; gastroduodenal disease.
OS Helicobacter pylori strain HP921023.
PN M09633220-A1.
PD 24-OCT-1996.
PF 19-APR-1996; AU0225.
PR 21-APR-1995; AU-002575.
PR 03-JUL-1995; AU-003931.
PR 16-JAN-1996; AU-007565.
PA (CSLC-) CSL LTD.
PA (UYNE-) UNIV NEW SOUTH WALES.
PI Doldge CV, Hocking DM, Lee A, Radcliff FJ, Webb EA;
DR WPI: 96-485735/48.
DR N-PSDB: T43046.
PT New Helicobacter antigens - used in vaccines for treatment or
PT prevention of Helicobacter infection, esp. in humans
PS Example 2; Page 60-61; 86pp; English.
CC An approx. 29 kDa protective antigen (W05197), designated H5.1, of
CC Helicobacter pylori is useful in the treatment or prevention of
CC gastroduodenal disease associated with H. pylori infection.

CC Protective antigens (W05194-98) were identified by screening a DNA
CC library of H. pylori with serum from mice immunised with H. felis
CC sonicate and isolating E. coli clones expressing H. pylori proteins
CC recognised by anti-H. felis antibodies. Cloned proteins were
CC grouped into families (see also W05194-98), and representative cDNA
CC clones of Families C, E, G, H and B (T43043-47) were sequenced.
CC Recombinant protective antigens can be produced for use in vaccines.
SQ Sequence 256 AA;

Query Match 3.5%; Score 91; DB 19; Length 256;
Best Local Similarity 27.3%; Pred. No. 4.26e+01;

Matches 15; Conservative 14; Mismatches 23; Indels 3; Gaps 2;

Db 168 gnveaylllddyysgndqg-lepdkkxayvykmaadmsrrtqeglaesyr 220

QY 251 GAGADFWLGGMLAGHSESGELIERGKKYLFYGSSEMAKMK-YAGVAEYR 304

RESULT 9
ID R08269 standard; protein; 265 AA.

AC R08269;

DT 07-MAR-1991 (first entry)

DE Mouse serine protease precursor, deduced from the DNA.

KM Leukemia elastase; pulmonary emphysema; rheumatoid arthritis;
OS Mus musculus.

PH key Location/Qualifiers
FT protein 29..265
/Label= serine protease.

PN J02268680-A.

PD 02-NOV-1990.

PF 07-APR-1989; 088798.

PR (TORA) TORAY IND INC.

PA WPI; 90-372006/50.

DR N-PSDB; Q06870.

PT Mouse model system - comprises mouse serine protease which shows
PS biological activity by polypeptide.

PS Claim 2; Page 1; 6pp; Japanese.

CC A mouse model system involving the serine protease can be used to
CC investigate the role of leukemia elastase in the development or
CC exacerbation of diseases such as pulmonary emphysema, rheumatoid
CC arthritis etc.

CC Sequence 265 AA;

Query Match 3.5%; Score 92; DB 2; Length 265;
Best Local Similarity 23.2%; Pred. No. 3.63e+01;

Matches 19; Conservative 30; Mismatches 26; Indels 7; Gaps 6;

Db 185 nvccllyrpgqgdcfgdsgpvlvcnmlvgqids-flr-ggcg-sglypdafapvafadw 241

QY 202 SVCT-TRKKTGVGVYQLSAVMECADAHGKHIIISDGGCSPGDVAKAGAGADPVM- 258

Db 242 Insirshndhllhpkdreg 263

QY 259 LCGMLAGHSESG-GELEIRDOCK 279

RESULT 10
ID P80335 standard; protein; 267 AA.

AC P80335;

DT 04-OCT-1990 (first entry)

DE Sequence of serine protease (SP) of human myeloid cell

KM Serine protease; myeloid cell; intravascular coagulation treatment;

OS Homo sapiens.

PH key Location/Qualifiers
FT peptide 1..29
PN M0806621-A.
PD 07-SEP-1988.
PE 26-FEB-1988; J00205.

PR 09-SEP-1987; JP-225540.
PA (TORA) Toray Ind Inc (AOXI/).

PI Aoki Y, Okano K, Naruto M, Shimizu H, Nakamura H;

DR WPI; 88-271161/38.

DR N-PSDB; N80315.

PT Serine protease of human myeloid cell origin -

PS Disclosure; Fig 6; 52pp; Japanese.

CC A SP of human myeloid cell origin is new. Also new are precursors of SP

CC having an N-terminal splittable or signal peptide; DNA sequences coding

CC for SP and precursors and transcription-controlling DNA sequence required

CC for expression of the gene. SP has antithrombotic activity, useful in

CC disseminated intravascular coagulation. By expression of protease gene

CC SP DNA selected and cloned from a library of human myeloid cell origin,

CC organism such as E.coli HMS-174 or HB-101 a yeast.

CC Sequence 267 AA;

Query Match 3.5%; Score 91; DB 2; Length 267;
Best Local Similarity 30.5%; Pred. No. 4.26e+01;

Matches 18; Conservative 19; Mismatches 17; Indels 5; Gaps 5;

Db 184 snvccllyrpgqgdcfgdsgpvlvcnmlvgqids-flr-ggcg-sglypdafapvafv 239

QY 201 GSVC-TR-KKTGVGVYQLSAVMECADAHGKHIIISDGGCSPGDVAKAGAGADPV 257

RESULT 11
ID R39819 standard; protein; 422 AA.

AC R39819;

DT 12-JUN-1994 (first entry)

DE ZmPK1 homologue protein kinase.

KM Kinase; homologue; differentiation; prolongation; proliferation;

OS Solanaceae.

PN J05161494-A.

PD 28-JUN-1993.

PF 13-DEC-1991; 330418.

PR (SUMO) SUMITOMO CHEM CO LTD.

PA WPI; 93-239258/30.

DR N-PSDB; Q47060.

PT ZmPK1 homologue protein kinase gene from solanaceae - used for

PT control of differentiation, prolongation and proliferation of

PS Claim 1; Page 5-6; 6pp; Japanese.

CC ZmPK1 homologue protein kinase or its gene is useful for the control

CC of differentiation, prolongation and proliferation of plants. The

CC gene can also be used to study agents which enhance or inhibit ZmPK1

CC homologue protein kinase.

SQ Sequence 422 AA;

Query Match 3.5%; Score 92; DB 7; Length 422;
Best Local Similarity 29.8%; Pred. No. 3.63e+01;

Matches 17; Conservative 19; Mismatches 19; Indels 2; Gaps 2;

Db 304 felitahpqnalmeyvnlamsdgydelldkklvgcslegv-tsjaiah-kci 358

QY 88 FSLFTANHKHYSLVQWQEFAGQNPDCLEHLAASSGSSPFDQLEQLEHLPQVKYI 144

RESULT 12
ID W26640 standard; protein; 520 AA.

AC W26640;

DT 29-JAN-1998 (first entry)

DE Arabidopsis thaliana ferulate-5-hydroxylase.

KM Ferulate-5-hydroxylase; lignin; delignification; pulp; paper;

KM feedstock; transgenic plant; alfalfa; Medicago; rice; Oryza;

KM maize; Zea mays; oilseed rape; Brassica; forage grass;

KM tobacco; Nicotiana; eucalyptus; pine; Pinus; spruce; Picea;

KM poplar; Populus; crop improvement; Arabidopsis thaliana.

PN M09723599-A2.
PD 03-JUL-1997.
PE 19-DEC-1996: U20094.
PR 14-MAR-1996: US-013388.
PR 22-DEC-1995: US-009119.
PA (DUPO) DU PONT DE NEMOURS & CO E. I.
PA (PORD) PURDUE RES. FOUND.
PI Chapelle C;
PI WPI: 97-351042/32.
DR N-PSDB: T90504-05.
PT Nucleic acid encoding plant ferulate-5-hydroxylase - useful for
PT altering quacetyl:syringyl lignin monomer ratio in plants, to
PT increase chemical delignification in paper and pulp making or to
PT increase digestibility of feedstocks.
PS Claim 1: Page 27-28; 43pp: English.
CC This polypeptide comprises Arabidopsis thaliana ferulate-5-
CC hydroxylase (F5H), a key enzyme in lignin biosynthesis that
CC catalyzes the conversion of ferulate to 5-hydroxyferulate. Its
CC amino acid sequence can be deduced from isolated cDNA and genomic
CC DNA clones (see T90504-05). A claimed chemtic gene that causes an
CC altered quacetyl:syringyl lignin monomer ratio in a transformed
CC plant comprises the F5H cDNA or genomic clone linked either in the
CC sense or antisense direction to at least one regulatory sequence.
CC The gene allows the alteration of a plant's lignin content,
CC especially to increase the syringyl lignin content. This renders
CC the lignin more susceptible to chemical delignification of use to
CC the paper and pulp industries. Alteration of the lignin content of
CC grasses increases the digestibility for animal feedstocks.
CC Suitable host plants include alfalfa, rice, maize, oilseed rape,
CC forage grasses, tobacco, and tree crops such eucalyptus, pine,
CC spruce and poplar.
SQ Sequence 520 AA;

[illegible][illegible]

RESULT	14
ID	W34892 Standard; Protein: 307 AA.
AC	W34892;
DT	03-MAR-1998 (first entry)
DE	Novel human phosphorylase kinase gamma subunit.
KM	phosphorylase kinase gamma subunit; replacement therapy;
KW	glycogen utilisation disorder; muscle glycogenosis; hypoglycaemia;
KV	gene therapy.
OS	Homo sapiens.
FH	Key
FT	Domain
FT	Location/Qualifiers
FT	1..300
FT	/note= "putatively contains the catalytic kinase domain"
FT	Domain
FT	300..385
FT	/note= "putatively contains 2 calmodulin binding domains required to activate the protein complex"
FT	Binding_site
FT	111
FT	/note= "putative substrate binding site"
FT	154
FT	/note= "putative substrate binding site"
FT	155
FT	/note= "important for catalytic activity"
FT	168..170
FT	/note= "important for catalytic activity"
FT	Active_site
FN	US5683910-A.
PD	04-NOV-1997.
PF	13-SEP-1996; 713828.
PR	13-SEP-1996; US-713828.
PA	(IMCY-) INCYTE PHARM. INC.
PI	Bardman O, Goli SK;
DR	WPI; 97-548985/50.
N-PDB:	T93887.
PT	New nucleic acid encoding human phosphorylase kinase gamma subunit - used for replacement, or gene, therapy of glycogen utilisation disorders, also for drug screening and diagnosis
PS	Claim 1; Fig 1; 25pp; English.
CC	The present sequence represents a novel human phosphorylase kinase gamma subunit, characterised as having homology to other phosphorylase kinase gamma subunits. Recombinant human phosphorylase gamma subunit is useful for replacement therapy in cases of glycogen utilisation disorders, e.g. muscle glycogenosis, causing muscular weakness and atrophy, or hypoglycaemia, causing reduced growth and kidney damage. Cells transformed with the DNA encoding the present sequence can also be used to screen for modulators of human phosphorylase kinase gamma subunit activity, i.e. potential therapeutic agents. The protein can also be used similarly in gene therapy, while antisense nucleic acid or ribozymes can be used to suppress mutant forms of the this gene. The protein can be used to raise specific antibodies, useful as immunoassay reagents. Fragments of the DNA encoding the present sequence can be used diagnostically, e.g. to detect or quantify gene expression, to isolate genes or related nucleic acid, to detect activation/induction of human phosphorylase kinase gamma subunit-expressing genes and for mapping the gene.
Sequence	307 AA:

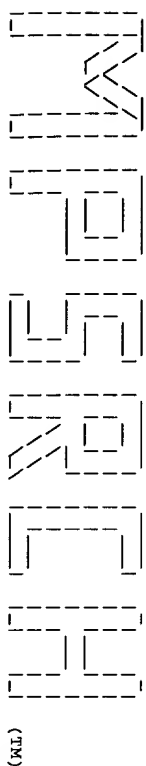
Query Match 3.4%; Score 89; DB 26; Length 307;
Best Local Similarity 30.6%; Pred. No. 5,86+01;

Matches 15; Conservative 14; Mismatches 17; Indels 3; Gaps 3;
 Db 37 yetasemllvleyaagglfdgcvarnd-afkexygrlmqlllegyh 84
 QY 285 YGMSSEMMMK-KTAGVAERASEKTYVEPPKG-DVEHTIDILGIR 331

RESULT 15
 ID W03516 standard; Protein; 359 AA.
 AC W03516.
 DT 11-MAR-1997 (first entry)
 DE Prostaglandin DP receptor.
 KW Prostaglandin DP receptor; human; prostanoid receptor; blood platelet;
 KM smooth muscle; nervous tissue; G protein-coupled receptor; modulator;
 KW mouse; prostaglandin-related disease; therapy.
 OS Homo sapiens.
 PN W09623066-A2.
 PD 01-AUG-1996.
 PF 23-JAN-1996; CA0047.
 PR 26-JAN-1995; US-378682.
 PA (MERI) MERCK FROSS CANADA INC.
 FI Abramovitz M, Bole Y, Welters K, Sawyer N, Slietz DM;
 DR WPI; 96-362690/36.
 DR N-PSDB; T37402.
 PT Human prostaglandin DP receptor and related DNA - used to identify
 PT receptor modulators to treat prostaglandin-related diseases.
 PS Claim 2, Page 41: 49pp. English.
 CC This sequence represents the human prostaglandin DP receptor. The DP
 CC receptor is the least ubiquitous and least abundant of the prostanoid
 CC receptors. The DP receptors are thought to be distributed mainly in
 CC blood platelets, smooth muscle of various tissues, and nervous tissue
 CC (including the central nervous system). However, many of the actions and
 CC distribution of the DP receptor is species dependent. This receptor is a
 CC G protein-coupled receptor which can specifically bind prostaglandin
 CC molecules. The DNA encoding this sequence was isolated using primers
 CC (see T37403 and T37404) based on the N-terminal sequence, and an internal
 CC sequence from purified mouse DP. The human prostaglandin DP receptor can
 CC be used to identify modulators of the receptor. The identified
 CC modulators can then be used to treat prostaglandin-related diseases, and
 CC for modulating the effects of prostaglandins on the DP receptor.
 SQ Sequence 359 AA;

Query Match 3.4%; Score 89; DB 20; Length 359;
 Best Local Similarity 24.7%; Pred. No. 5.86e+01;
 Matches 22; Conservative 26; Mismatches 34; Indels 7; Gaps 7;
 Db 212 latvclnlgamrnyamhrrl-qtrhpsc-trdcapradgreasppleelnhlllall 269
 QY 81 MAKVLCRFSLEFVAH-KHSLVQWQEFAGQNPDCLEHLAASSGTSSDFEQLEQILE-AI 138
 Db 270 mtvlfmc-slpvlyrayyga-f-kdvkek 296
 QY 139 POKY-ICLDVANGYSEHFVEFVKDKRR 166

Search completed: Mon Jun 15 15:58:12 1998
 Job time : 98 secs.



Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_p protein - protein database search, using Smith-Waterman algorithm
Run on: Mon Jun 15 15:58:31 1998; MasPar time 16.06 Seconds
Tabular output not generated. 832.618 Million cell updates/sec

Title: >US-08-774-169-1
Description: (1-366) from US08774169.pep
Perfect Score: 2632
Sequence: 1 MTCCPALRFTATPRLSAMP.....RRTFIRVYQVNPISFAC 366

Scoring table: PAM 150
Gap 11

Searched: 120441 seqs, 36531193 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir56
1:pir1 2:pir2 3:pir3 4:pir4 5:nr13d

Statistics: Mean 47.551; Variance 98.999; scale 0.480

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	2052	78.0	345	2	B32902	GMP reductase (EC 1.6	0.00e+00
2	1736	66.0	347	2	H64732	GMP reductase (EC 1.6	0.00e+00
3	550	20.9	488	1	DEBSMP	IMP dehydrogenase (EC	2.81e-78
4	532	20.2	481	2	E64673	inosine 5'-monophosph	7.74e-75
5	529	20.1	326	2	C70015	GMP reductase homolog	2.89e-74
6	529	20.1	493	2	JC4372	IMP dehydrogenase (EC	2.89e-74
7	497	18.9	404	2	E70218	IMP dehydrogenase (EC	3.52e-68
8	491	18.7	496	2	G64501	IMP dehydrogenase (EC	4.83e-67
9	476	18.1	488	1	DEECIP	IMP dehydrogenase (EC	3.31e-64
10	474	18.0	488	2	S23226	IMP dehydrogenase (EC	7.89e-64
11	474	18.0	488	1	H64055	IMP dehydrogenase (EC	7.89e-64
12	466	17.7	484	2	B69056	inosine 5'-monophosph	2.54e-62
13	455	17.3	485	2	JC4998	IMP dehydrogenase (EC	2.98e-62
14	432	16.4	327	2	F64626	GMP reductase - Helic	6.06e-56
15	385	14.6	529	2	S72823	IMP dehydrogenase (EC	3.17e-47
16	364	13.8	503	2	UC4999	IMP dehydrogenase (EC	2.25e-43
17	354	13.4	514	2	J70565	IMP dehydrogenase (EC	1.50e-41
18	354	13.4	514	2	A35566	IMP dehydrogenase (EC	1.50e-41
19	349	13.3	514	2	A31997	IMP dehydrogenase (EC	1.22e-40
20	347	13.2	514	2	B31997	IMP dehydrogenase (EC	2.81e-40
21	345	13.1	537	2	SA1064	IMP dehydrogenase (EC	6.47e-40
22	343	13.0	523	2	S59402	probable IMP dehydrog	1.49e-39
23	343	13.0	524	2	S50890	probable IMP dehydrog	1.49e-39

24	339	12.9	523	2	S48997	probable IMP dehydrog	7.89e-39
25	336	12.8	512	2	A53407	IMP dehydrogenase (EC	2.79e-38
26	328	12.5	537	2	S59508	IMP dehydrogenase (EC	7.60e-37
27	310	11.8	514	2	A38668	IMP dehydrogenase (EC	1.27e-33
28	287	7.7	403	2	S53477	probable IMP dehydrog	4.09e-15
29	201	7.6	387	2	S75050	IMP dehydrogenase cha	5.94e-15
30	163	6.2	157	2	S78078	hypothetical protein	5.52e-09
31	125	4.7	92	5	1AK52	IMP dehydrogenase (EC	1.59e-03
32	104	4.0	174	2	S59994	AMP deaminase (EC 3.5	7.71e-01
33	104	4.0	256	2	S59995	AMP deaminase (EC 3.5	7.71e-01
34	104	4.0	605	2	S59996	AMP deaminase (EC 3.5	7.71e-01
35	104	4.0	760	2	A44313	AMP deaminase (EC 3.5	7.71e-01
36	103	3.9	342	2	S51402	probable membrane pro	1.02e+00
37	103	3.9	508	1	A56008	probable serine-type	1.02e+00
38	102	3.9	511	2	A69369	glutamate synthase (g	1.34e+00
39	102	3.9	622	2	E69006	glutamate synthase (N	1.34e+00
40	101	3.8	168	2	A34952	CS3 fibrillar protein	1.76e+00
41	100	3.8	357	2	E64374	hypothetical protein	2.30e+00
42	98	3.7	168	2	S07904	CS3 pilin precursor	3.93e+00
43	98	3.7	320	5	1GDH8	glycerate dehydrogena	3.93e+00
44	98	3.7	320	5	1GDH8	glycerate dehydrogena	3.93e+00
45	98	3.7	383	2	E69099	pyruvate oxidoreducta	3.93e+00

ALIGNMENTS

RESULT	1	ALIGNMENTS
ENTRY	B32902	#type complete
TITLE	GMP reductase (EC 1.6.6.8) - human	
ALTERNATE_NAMES	guanosine monophosphate reductase	
ORGANISM	#formal_name Homo sapiens #common_name man	
DATE	20-Dec-1989 #sequence_revision 20-Dec-1989 #text_change 10-Sep-1997	
ACCESSIONS	B32902; A32902; A55285	
REFERENCE	A32902	
#authors	Kanno, H.; Huang, I.Y.; Kan, Y.W.; Yoshida, A.	
#journal	Cell (1989) 58:595-606	
#title	Two structural genes on different chromosomes are required for encoding the major subunit of human red cell glucose-6-phosphate dehydrogenase.	
#cross-references	glucose-6-phosphate dehydrogenase. PMID:89336791	
#accession	B32902	
#molecule_type	mRNA	
#residues	1-345 #label KAN	
#cross-references	GB:M2470; NID:g182866; PID:g182867; GB:M27958	
#accession	A32902	
#molecule_type	protein	
#residues	43-35; 'GASGDLAK' #label KA2	
#note	direct peptide sequencing of a single fragment of a chimeric protein with sequence from glucose-6-phosphate 1-dehydrogenase (A40309) and this protein; the chimeric product was later shown to be an artifact of purification and not a natural component of erythrocytes	
REFERENCE	A55285	
#authors	Kondoh, T.; Kanno, H.; Chang, L.; Yoshida, A.	
#journal	Hum. Genet. (1991) 88:219-224	
#title	Genomic structure and expression of human guanosine monophosphate reductase.	
#accession	A55285	
#status	preliminary; translation not shown	
#molecule_type	DNA	
#residues	1-233, 'A', '235-255', 'F', '257-345' #label KON	
#cross-references	GB:S73035	
REFERENCE	A55176	
#authors	Henikoff, S.; Smith, J.M.	
#journal	Cell (1989) 58:1021-1022	
#title	The human mRNA that provides the N-terminus of chimeric G6PD encodes GMP reductase.	
REFERENCE	A55177	
#contents	annotation	
REFERENCE	Yoshida, A.; Kan, Y.W.	
#authors	Cell (1990) 62:11-12	


```

#authors
Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Biolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, V.; Brouillet, S.;
Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Connetton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesthofs, A.;
Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
Fabre, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Gilh,
S.Y.; Glaser, P.; Goffeau, A.; Gough, S.; Grandi, G.;
Giles, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holapfel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kasahara, Y.; Klier-Blanchard, M.; Klein, C.; Kobayashi,
I.; Koetler, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Patro,
V.; Pohl, T.M.; Portelle, D.; Portelle, S.; Prescott,
A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serro, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vasarotti, A.; Vialti, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Welleneger, T.;
Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasunoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumbstein, E.;
Yoshikawa, H.; Danchin, A.
Nature (1997) 390:249-256
The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.

#journal
#title
#accession
#status
#molecule_type
#residues
#experimental_source
#length
#molecular_weight
#checks
#yund
#gene
#genetics
SUMMARY
Query Match
Best Local Similarity
Matches
118:
Conservative
79:
Mismatch
116:
Indels
21:
Gaps
17:
5 FDVEDIOLIPAKCIVNSRSEC-T-S-V-RLGHTFK-LPVYANMOTIIDEKLAISLAE
28 LDFEDVLLRPRSRSLKRSFVLDLRSFNSKOTYSGVPIAANMOTVGFENAKYLC
60 NGIYYVNR-FEPEIRDFIDMNRGL-FSSISVGVKDEYEFVROLAENILPEVYTI
88 FSLETAHKKHYSLVQOEF-A-GQNPDLLEHLAASSGTGSDFEOLLEAIPOVKYICL
118 DIANGHNAVEMQHKLHPDSFVLAGNVTPEAVRELENAADATKVGIGKXCIT
147 DVANGISEHFEVYKDKRPPQHTIMAGNVTEMEVELLSADITIKVIGSGVCTT
178 KIKTGTGGMOLALRWCAKA-S-K-PIADGIRTHGDIASIRFATMAGISLFA
207 RKKTGVYR--QLSAVWECADNAHGLKHTISDGCSCPGVAAKAFAGADFYVLGMLA
235 GHEESPQITLTKDKLYKEFGSASEPP--K--G-EK-KNVEGKMMHVAHKGSIKDTLI
265 GHSGSGELLERDCKKYLFGMSSEMAKKYAGVALEYRASEGKTEVPEPKGVERTIR
288 EMEDDLSSISYAGGTLMNI-RNVYIVYVKNST 320

#authors
Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Biolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, V.; Brouillet, S.;
Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Connetton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesthofs, A.;
Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
Fabre, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Gilh,
S.Y.; Glaser, P.; Goffeau, A.; Gough, S.; Grandi, G.;
Giles, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holapfel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kasahara, Y.; Klier-Blanchard, M.; Klein, C.; Kobayashi,
I.; Koetler, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Patro,
V.; Pohl, T.M.; Portelle, D.; Portelle, S.; Prescott,
A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serro, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vasarotti, A.; Vialti, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Welleneger, T.;
Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasunoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumbstein, E.;
Yoshikawa, H.; Danchin, A.
Nature (1997) 390:249-256
The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.

#journal
#title
#accession
#status
#molecule_type
#residues
#experimental_source
#length
#molecular_weight
#checks
#yund
#gene
#genetics
SUMMARY
Query Match
Best Local Similarity
Matches
99:
Conservative
58:
Mismatch
80:
Indels
10:
Gaps
5:
225 VAAAVGTSPTFERALEFAGADA--TYVDTHGHSAGYLRKIAIEIRAHFPRRTLAGN
117 LAASSGSGSDFEOLLEAIPOVKYICLDVANGYSEHFEVVKDKRFPQHTIMAGN
283 IATAGARALYDAGVDYVYKIGSGICTRRVAVGVPQVNTIYDAAANAREYGTIIA
177 VVGEMVEELLISGADIKKIGIGSVCTTKTKTGVPOLSVMCADAAHGLKHTIS
343 DGGIKYSGDIYKALAGAGNAVMLGSEFAGTDEAPGETEYQGRKFTYRGMGSIAAMKKG
237 DGGCSCPGDVAAKAFGACADVYMLGMLAGHSESGELIENDCKKYLFGYMSSEMAKK-
403 SSDRYFGGSVNEANKLVPEIGRVAAYKGAASDIYFOMLGIRSGMGYVAGDIQELHEN
296 ---Y-AGVAAE-YR-ASEGKTVEVPFGVDERTIRDLIGIRSTCTYVGAARKLERR
463 AOPVEMS 469
349 TTFIRVT 355

RESULT
ENTRY
TITLE
ALTERNATE_NAMES
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.;
Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.;
Dodson, R.; Hickey, E.K.; Gwinn, M.; Dougherty, B.; Tomb,
J.F.; Riedelmann, R.D.; Richardson, D.; Peterson, J.;
Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.;

```


#title Nucleotide sequence of the guaB locus encoding IMP
#cross-references MUID:85215547
#accession A93555
##molecule-type DNA
##residues 'MOSVTCIMPROYLTTLVELP', 1-205, 'A', 207-488 ##label TIE
##cross-references GB:X02209; NID:g41626; PID:g41627
##experimental source strain K12
REFERENCE
#authors Thomas, M.S.; Drabble, W.T.
#journal Gene (1985) 36:45-53
#title Nucleotide sequence and organisation of the gua promoter
region of *Escherichia coli*.
#cross-references MUID:8605659
#accession A91529
##molecule-type DNA
##residues 1-96 ##label THO
REFERENCE
#authors Testa-Selase, F.; Drabble, W.T.
#journal Mol. Gen. Genet. (1992) 231:256-264
#title Regulation of the gua operon of *Escherichia coli* by the DnaA
protein.
#cross-references MUID:92140365
#accession S20017
##status not compared with conceptual translation
##molecule-type DNA
##residues 1-90 ##label TES
COMMENT The active enzyme, a tetramer of identical chains, catalyzes the
conversion of IMP to XMP in the presence of NAD; this reaction is
the first unique step to GMP biosynthesis.
GENETICS
#gene guaB
#map-position 54 min
CLASSIFICATION #superfamily IMP dehydrogenase
KEYWORDS GMP biosynthesis; NAD; oxidoreductase; purine nucleotide
biosynthesis
FEATURE
305 #active-site Cys #status predicted
SUMMARY #length 488 #molecular-weight 52022 #checksum 7566
Query Match 18.1%; Score 476; DB 1; Length 488;
Best Local Similarity 35.9%; Pred. No. 3,316-64;
Matches 89; Conservative 69; Mismatches 79; Indels 11; Gaps 8;
Db 219 RVGAAGAGAGNERDALVAGVDV--LL-IDSSHSEGVLDQRTRETRAKYPPDLQITGG 276
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 116 HLAASGSGSDFEQQLLEAIPOVKYICLDVANGYSEHVEFVKDYKRFPQHTIMAG 175
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 277 NVATAAGABALAEAGCAVKYIGIGISICTRIYGVGPQITAVADAVEALEGCIPIVI 336
||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 176 NVVTGEMVEELILSGADILKIVGIGPSCVTRKKTGVGPQLSAVMCADNAHGLKGHIT 235
||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 337 ADGIRFSGDIKAIAGASAVMVGSLAGTEESPELELYOGRSYKSYRGMSLGAWSK 396
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 236 SDGCSGCPDVAKAFAGADFMVLMGLAGHSESGELIERDKKYLKFGMS--EKAM 293
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 397 GSSDRYFSDNAAKLVPEGIEGRVAYKGRKEITHQMGSLRS-CMGLTSCGTTDELRT 455
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 294 --KKY-AGVAYEYR-ASEGKTVEYFPEKGDVEHTIRDLIGIRSYCT-YVGAALKLELSR 347
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 456 KAEFVRIS 463
:::|:::|:::|
QY 348 RTTIRIVT 355

REFERENCE S23225
#authors Anderegg, U.; Schnuck, W.H.; Asperger, O.; Kleber, H.P.
#submission submitted to the EMBL Data Library, June 1992
#accession S23226
##molecule-type DNA
##residues 1-488 ##label AND
##cross-references EMBL:X66859; NID:g38718; PID:g38720
CLASSIFICATION #superfamily IMP dehydrogenase
KEYWORDS oxidoreductase
SUMMARY #length 488 #molecular-weight 51530 #checksum 7816
Query Match 18.0%; Score 474; DB 2; Length 488;
Best Local Similarity 38.9%; Pred. No. 7,896-64;
Matches 98; Conservative 69; Mismatches 68; Indels 17; Gaps 11;
Db 218 RVGAAGTADTPSRVEALVEAGVDV--IYVDAHGSAGVIERVRYKONFPQVYIG 275
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 116 HLAASGSGSDFEQQLLEAIPOVKYICLDVANGYSEHVEFVKDYKRFPQHTIMAG 175
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 276 NIATGDALALDGAANVYIGPISCTTRIVAGTGMQISAI-DSV-AS-ALKQDIP 332
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 176 NVVTGEMVEELILSGADILKIVGIGPSCVTRKKTGVGPQLSAVMCADNAHGLKGHIT 234
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 333 LIADGIFPSGDMAKATGAGASTIMVGSLLAGTEAPGEVFEFGRRYKAYRGMSIGAM 392
||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 235 -ISDGCSCPDYAKAGAGADFMVLMGLAGHSESGELIERDKKYLKFGMSSEMAM 293
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 393 AGRTSADRYFODSKAGAEKLVPEGIEGRVYKPGMGNIVHQMVGSLRSGYTSAYIE 452
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 294 KKVAGVAYEYR-AS-EG-K--T--VE--VPKGDVEHTIRDLIGIRSYCTYGAALKR 343
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 453 DLROMAKFVKIT 464
:::|:::|:::|
QY 344 ELSRRTTIRIVT 355

RESULT 11
ENTRY H64055 #type complete
TITLE IMP dehydrogenase (EC 1.1.1.205) - Haemophilus influenzae
ALTERNATE_NAMES inosine-5'-monophosphate dehydrogenase
ORGANISM #formal name Haemophilus influenzae
DATE 18-Aug-1995 #sequence-revision 04-Oct-1996 #text-change
10-Oct-1997
ACCESSIONS H64055
REFERENCE A64000
#authors Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.;
Kirkness, E.F.; Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.;
Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.;
Fitzhugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;
Shirley, R.; Liu, L.T.; Glodek, A.; Kelley, J.M.; Weldman,
J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton,
M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek,
J.L.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann,
J.L.; Geoghegan, N.S.M.; Gnehm, C.L.; McDonald, L.A.;
Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.;
Science (1995) 269:496-512
#journal Whole-genome random sequencing and assembly of Haemophilus
#title influenzae Rd.
#cross-references MUID:95350630
#accession H64055
##status nucleic acid sequence not shown; translation not shown
##molecule-type DNA
##residues 1-488 ##label TIGR
##cross-references GB:U32708; GB:L42023; NID:g1573183; PID:g1573185;
TIGR:HI0221
COMMENT The active enzyme, a tetramer of identical chains, catalyzes the
conversion of IMP to XMP in the presence of NAD; this reaction is
the first unique step to GMP biosynthesis.
GENETICS
#gene guaB
CLASSIFICATION #superfamily IMP dehydrogenase
KEYWORDS GMP biosynthesis; homotetramer; NAD; oxidoreductase; purine

FEATURE nucleotide biosynthesis

305 #active_site Cys #status predicted

SUMMARY #length 488 #molecular-weight 51980 #checksum 8585

Query Match 18.0%; Score 474; DB 1; Length 488;

Best Local Similarity 36.9%; Pred. No. 7,896-64;

Matches 92; Conservative 68; Mismatches 76; Indels 13; Gaps 10;

Db 220 RVGAAGAGAGNERIDALVAGVDY-LL-IDSSHGSEGLVLOVRENRATYPLPIYAG 277

QY 116 HLAASGSGSDFEQLEILAIPOVKYICLDVANGYSEHFVEYKDKRFPQHTIAG 175

Db 278 NVATAEAGALADAGASAVKVGIGPSICTRIYGVGPQITAIADAALAKD-RGIPV 336

QY 176 NVATGEVIELISGADIKVIGIGSVCTTRKKTGVGPQLSAVMECADAHGK-GHI 234

Db 337 IADGIRPSGDIARAIAAGASCVAGSMFAGTEAPGTEIELYQGRAPKSYMGSLGAMA 396

QY 235 ISDGSCSPGVAFAGAFGADFMGLMAGHSESGELIRDSKRYKLFYGMSS--EMA 292

Db 397 KGSSDRYFQSNADKLVEGRIPIRYKYLKELIHQOMGLRS-CMGLTGCAITIDELR 455

QY 293 M---KKT-AGGVAEYR-ASEGKIVEPFGVDEHTIRDLGIRSTCT-YVGAALKLELS 346

Db 456 TRAEFVRIS 464

QY 347 RRTTFIRYT 355

RESULT 12

ENTRY B69056 #type complete

TITLE Inosine-5'-monophosphate dehydrogenase - Methanobacterium

ORGANISM thermoaototrophicum (strain Delta H)

DATE #formal_name Methanobacterium thermoaototrophicum

05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change

26-Feb-1998

ACCESSION B69056

REFERENCE A69000

#authors Smith, D.R.; Doucette-Stamm, L.A.; Delonghey, C.; Lee, H.; Dubois, J.; Alcedge, T.; Bashirzadeh, R.; Blakely, D.; Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.; Lum, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaltre, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jivani, N.; Caruso, A.; Bush, D.; Safer, H.; Patwell, A.; Prabhakar, S.; McDougall, S.; Shiner, G.; Goyal, A.; Pietrkowski, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Moelling, J.; Reeve, J.N.

#journal J. Bacteriol. (1997) 179:7135-7155

#title Complete genome sequence of Methanobacterium thermoaototrophicum Delta H: functional analysis and comparative genomics.

#cross-references M01D:96037514

#accession B69056

#status preliminary; nucleic acid sequence not shown; translation not shown

##molecule_type DNA

##residues 1-484 #label MTH

##cross-references GB:AE000666

##experimental_source strain Delta H

GENETICS

CLASSIFICATION MTH142

SUMMARY #superfamily IMP dehydrogenase

#length 484 #molecular-weight 52206 #checksum 6744

Query Match 17.7%; Score 466; DB 2; Length 484;

Best Local Similarity 36.1%; Pred. No. 2,546-62;

Matches 97; Conservative 68; Mismatches 88; Indels 16; Gaps 9;

Db 203 ILKRRYPNARSDESGYRVAAAGFPDLR-ARALDE-AGADVLADISAGHNKLVKS 260

QY 100 LVQWQEFAGNDPLEHLAASSGTGSDFEQLEILAIPOVKYICLDVANGYSEHFVEF 159

Db 261 AGAMKEIEADLIY-GNIATREAAEDLIADVDGLKVGIGPSMCTTRIAGVPOLTA 319

QY 160 VKDYRRFPQHTIAGNVVATGEVIELISGADIKVIGISVCTTRKKTGVGPQLSA 219

Db 320 IAEVADVAEYGVFVLDGIRISGDIARAIAVAGDCVMGLNLAGTYEAPGVYVNR 379

QY 220 VMECADAHGKGIHIDGSCSGDVAKAFAGADFMGLMAGHSESGELIENDGK 279

Db 380 KYKQRMGSLGAMTGGISGTDTRFQEPGRHKHTVPEGVGPYRGTVSEVLFOL 439

QY 280 KYKLFYGMSSAMKKTAGVAAE-Y-RASG--K-T-----VE--VFGVDVHTIRDI 326

Db 440 IGLRASMGYCGAANLSEMKERATVRIT 468

QY 327 LGIRSTCTYVGAALKLELSRRTTFIYV 355

RESULT 13

ENTRY JC4998 #type complete

TITLE IMP dehydrogenase (EC 1.1.1.205) - Pyrococcus furiosus

ORGANISM #formal_name Pyrococcus furiosus

DATE 12-Dec-1996 #sequence_revision 21-Jan-1997 #text_change

08-Sep-1997

ACCESSION JC4998

REFERENCE JC4998

#authors Collart, F.R.; Osipuk, J.; Trent, J.; Olsen, G.J.; Huberman, E.

#journal Gene (1996) 174:209-216

#title Cloning, characterization and sequence comparison of the gene coding for IMP dehydrogenase from Pyrococcus furiosus.

#accession JC4998

#status preliminary

##molecule_type DNA

##residues 1-485 #label COL

##cross-references GB:U08814; NID:9595286; PID:9595287

GENETICS

#gene quab

COMPLEX homotetramer

FUNCTION #description provides precursors for DNA and RNA biosynthesis; it catalyzes the conversion of IMP to XMP in the presence of NAD; this reaction is the first unique step to GMP biosynthesis

CLASSIFICATION #superfamily IMP dehydrogenase

KEYWORDS GMP biosynthesis; NAD; oxidoreductase; purine nucleotide

SUMMARY #length 485 #molecular-weight 52900 #checksum 6440

Query Match 17.3%; Score 455; DB 2; Length 485;

Best Local Similarity 36.6%; Pred. No. 2,986-60;

Matches 89; Conservative 58; Mismatches 83; Indels 13; Gaps 8;

Db 222 AAAYSPD-LRAIELDRAGVIVDTAHHNLKAIKANKEMOKVSADFTV-GNIANP 279

QY 122 GTGSDPEQLEQILE-AIPQYKICLDVANGYSEHFVEYKDKRFPQHTIAGNVYTG 180

Db 280 KAVVDLTF--ADAVKVGIGPSICTRIYAGVGPQITAIAMVADRAQEGLYIADGI 337

QY 181 EMVELISGADIKVIGIGSVCTTRKKTGVGPQLSAVMECADAHGKGIHIDSGC 240

Db 338 KYSGDIKAIAGADAVMLGNLAGTKEAPGKEVINGRKYKOVYRGMSLGAMKGAER 397

QY 241 SCPEADVAKARAGADFMGLMAGHSESGELIERGKRYKLFYGMSS--EMAKK---K 295

Db 398 YVGGYKTRKTFPEGVGVYPRYGVSEVLYOLVGLKAGMGVGAARINKELEKEGEFV 457

QY 296 Y-AGGVAEYR-ASEGKIVEPFGVDEHTIRDLGIRSTCTYVGAALKLELSRRTTFI 352

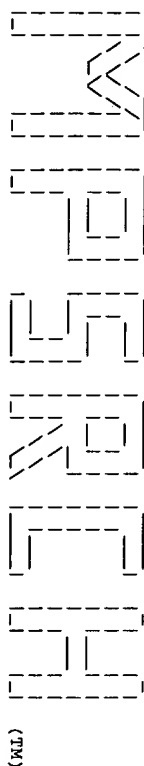
Db 458 IIT 460

QY 353 RVT 355

RESULT	ENTRY	TITLE	ORGANISM	DATE	ACCESSIONS	REFERENCE	#authors
14	F64626	#type complete	GMP reductase Helicobacter pylori (strain 26655)	09-Aug-1997	#formal_name Helicobacter pylori	09-Aug-1997	#sequence_revision 09-Aug-1997
	F64626			10-Oct-1997			
	F64626						
	A64520						
	Tomb, J.F., White, O., Kierlavage, A.R., Clayton, R.A., Sutton, G.G., Fleischmann, R.D., Ketchum, K.A., Klenk, H.P., Gill, S., Dougherty, B.A., Nelson, K., Quackenbush, J.P., Zhou, L., Kirchner, E.F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khakh, H.G., Glodek, A., McKenney, K., Fitzgerald, L.M., Lee, N., Adams, M.D., Hickey, E.K., Berg, D.E., Goeckye, J.D., Utterback, T.R., Peterson, J.D., Kelley, J.M., Cotton, M.D., Weidman, J.M., Fujii, C., Bowman, C., Matthey, L., Wallin, E., Hayes, W.S., Borodovsky, M., Karp, P.D., Smith, H.O., Fraser, C.M., Venter, J.C.						
	Nature (1997) 388:339-547						
	The complete genome sequence of the gastric pathogen Helicobacter pylori.						
	#accession F64626						
	#status						
	##molecule_type DNA						
	##residues 1-327						
	##cross-references GB:AE00596; GB:AE00511; MID:g2313982; PID:g2313987; TIGR:HP0854						
	##length 327						
	##molecular_weight 36038						
	##checksum 7430						
	Query Match						
	Best Local Similarity 32.3%;						
	Matches 108; Conservative 79; Mismatches 126; Indels 21; Gaps 15;						
	6 FDYEDVOLINCKIVNSRSEC-D-TYI-L-V-GKHAK-R-MIYAPAMQIINESIAPAE 60						
	28 LDFEDVILRRKSTLSRSSEVDLIRSEFRNSQYTSYPIAAMDVYGFEEAKVICK 87						
	61 NGYFYIMHR-FDGAARIPFYKKKKR-QWISSISYVKKEECLFYEELAKQGLADPYTI 118						
	88 FSLFTANHKHISLVQNEFAGQNPDCLEHNA-SSTGSSDPEQLEQLIEALPYKTYICL 146						
	119 DIAHGSNSVIEIMQRIKTHLPETVYIAGNVGTPEAVRELENAGADATKVGIGPKVCIT 178						
	147 DVANGYSEHFVEFYKDVKKRFPQHTIMAGNVYTGVEWEELISGADILKVGIGSPVCYT 206						
	179 KIKGFGTGMOLALRMCAKAR-K-PIADGGIRTGIDYKIRGATVVMGSLFA 235						
	207 RKKGVYV-P-QLSAVECDAAHGKIHGISDSCGPDVAKKFGAGADFVMGGMIA 264						
	236 GHESSSETKIEINGIAWKEFGSASEF--OK--G-EK-KNIEGKIMVIOHGSIKDTLV 288						
	265 GHSSEGELLEROKKTKLFYGSSEMAKKYAGVAEFAEASEGKTVEVPRKGDVEHTIR 324						
	289 EMHODLQSSISYAGRDLEAI-RKYDVYIVKSI 321						
	325 DILGIRSTCTYVGAAKLKLRSRTTFIRVTVQV 358						
15	S72823	#type complete	IMP dehydrogenase (EC 1.1.1.205) guab2 - Mycobacterium lepreae	19-Mar-1997	#formal_name Mycobacterium lepreae	19-Mar-1997	#sequence_revision 25-Apr-1997
	S72823			08-Sep-1997			
	S72820						
	Smith, D.R.; Robison, K.						
	submitted to the EMBL Data Library, November 1993						
	Mycobacterium lepreae cosmid B1620.						

	#accession	S72823
	##molecule_type	DNA
	##residues	1-529 ##label
	##cross-references	EMBL:000015; NID:q466931; PID:q466944
	GENETICS	
	gene	guab2
	CLASSIFICATION	superfamily IMP dehydrogenase
	KEYWORDS	GMP biosynthesis; NMD; oxidoreductase; purine nucleotide biosynthesis
	SUMMARY	#length 529 #molecular-weight 54814 #checksum 2745
	Query Match	14.6%; Score 385; DB 2; Length 529;
	Best Local Similarity	35.7%; Pred. No. 3,17e-47;
	Matches	91; Conservative 69; Mismatches 75; Indels 20; Gaps
Db	25	VGAALVCGGDAMVRAMLVDAQDV-LI-VDTAHNRRLVLDYGLKVEIGDRVYIG 312
Qy	117	LAASGTGSSDFEDELQLEILAIPQYKICLDVANGYSEHFEKVDKRRFPQHT-IMAG 175
Db	313	NVATRSAAALVYENGADAVKVGSPGSTCTTRVAVGAGAPQITALE-AVAACGPACVPV 371
Qy	176	NVATGVEMVEELISGLADITIKVIGISPGVCTTRKKTGVGPOLSKVMPCADAAHGLK-KHI 234
Db	372	IADGGLQYSGDIAKALAAGASTYMLGSLAGTAEPPELLFVNGKOPKPSYRGMSGLAMQ 431
Qy	235	ISDGCCSCPGDVAAAFAGAGAFVWGLGMLAHGSHSSGSELLEROCKKTKLEYGMS--EM- 291
Db	432	GRGDGKYSKQRIRPADDALSDKLVPRGIEGRVPRFQPLSSVTHQLVGLRLPAMGYTSP 491
Qy	292	AM---K---K--Y-AG-GVAEYR-ASEGKTVEVPEFKGDVDEHTIIDIGIRSTCTYVGA 340
Db	492	TIEVL-OOAFVRIT 505
Qy	341	KLKELSRKTTIRVT 355

Search completed: Mon Jun 15 15:59:38 1998
Job time : 67 secs.



(TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_dp protein - protein database search, using Smith-Waterman algorithm
Run on: Mon Jun 15 15:59:54 1998; Maspar time 10.95 Seconds
Tabular output not generated. 838.649 Million cell updates/sec

Title: >US-08-774-169-1
Description: (1-366) from US08774169.pap
Perfect Score: 2632
Sequence: 1 MTCCLPALRFIATPRLSAMP.....RTTFIRVYQGVNPIFSEAC 366

Scoring table:
PAM 150
Gap 11

Searched: 69111 seqs, 25083644 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot35
1:swiss1

Statistics: Mean 49.471; Variance 81.711; scale 0.605

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	2048	77.8	345	1	GUAC_HUMAN	GMP REDUCTASE (EC 1.6. 0.00e+00	
2	1850	70.3	356	1	GUAC_ASCSU	GMP REDUCTASE (EC 1.6. 0.00e+00	
3	1709	64.9	346	1	GUAC_ECOLI	GMP REDUCTASE (EC 1.6. 0.00e+00	
4	550	20.9	513	1	IMDH_BACSU	IMDH_BACSU	
5	532	20.1	481	1	IMDH_HELPY	IMDH_HELPY	
6	529	20.2	493	1	IMDH_STRPY	IMDH_STRPY	
7	497	18.9	404	1	IMDH_BORBU	IMDH_BORBU	
8	491	18.7	496	1	IMDH_METUA	IMDH_METUA	
9	476	18.1	488	1	IMDH_ECOLI	IMDH_ECOLI	
10	474	18.0	488	1	IMDH_ACICA	IMDH_ACICA	
11	474	18.0	488	1	IMDH_HAEIN	IMDH_HAEIN	
12	455	17.3	485	1	IMDH_PYRPU	IMDH_PYRPU	
13	393	14.9	529	1	IMDH_MYCTU	IMDH_MYCTU	
14	385	14.6	529	1	IMDH_MYCTU	IMDH_MYCTU	
15	367	13.8	501	1	IMH3_CANAL	IMH3_CANAL	
16	364	13.8	503	1	IMH3_CANAL	IMH3_CANAL	
17	356	13.5	454	1	IMDH_PNECA	IMDH_PNECA	
18	354	13.4	514	1	IMDH_HUMAN	IMDH_HUMAN	
19	354	13.4	514	1	IMDH_MOUSE	IMDH_MOUSE	
20	349	13.3	514	1	IMDH_HUMAN	IMDH_HUMAN	
21	347	13.2	514	1	IMDH_MESAU	IMDH_MESAU	
22	345	13.1	537	1	IMDH_DROME	IMDH_DROME	
23	343	13.0	523	1	IMH2_YEAST	IMH2_YEAST	

24	343	13.0	524	1	IMH3_YEAST	PROBABLE INOSINE-5'-MO	8.56e-50
25	339	12.9	523	1	IMH1_YEAST	PROBABLE INOSINE-5'-MO	6.88e-49
26	336	12.8	512	1	IMDH_TRYBG	IMDH_TRYBG	3.28e-48
27	323	12.3	514	1	IMDH_MOUSE	IMDH_MOUSE	2.76e-45
28	318	12.1	514	1	IMDH_LEIDO	IMDH_LEIDO	3.63e-44
29	272	10.3	503	1	IMDH_TRLFP	IMDH_TRLFP	5.05e-34
30	202	7.7	403	1	IMH4_YEAST	IMH4_YEAST	2.52e-19
31	198	7.5	141	1	YOD1_MYCTU	YOD1_MYCTU	1.60e-18
32	107	4.1	375	1	YOD1_MYCTU	YOD1_MYCTU	3.33e-02
33	104	4.0	760	1	AMH2_HUMAN	AMH2_HUMAN	9.19e-02
34	103	3.9	508	1	YB9_YEAST	YB9_YEAST	1.28e-01
35	100	3.8	357	1	Y597_METUA	Y597_METUA	3.43e-01
36	97	3.7	92	1	DMPK_PSEPU	DMPK_PSEPU	8.96e-01
37	98	3.7	168	1	FMS3_ECOLI	FMS3_ECOLI	6.52e-01
38	98	3.7	321	1	DH93_HYPME	DH93_HYPME	6.52e-01
39	97	3.7	375	1	Y09J_MYCTU	Y09J_MYCTU	8.96e-01
40	95	3.6	199	1	IL11_MOUSE	IL11_MOUSE	1.68e+00
41	95	3.6	205	1	FIXJ_BRAJA	FIXJ_BRAJA	1.68e+00
42	95	3.6	289	1	SYN3_HUMAN	SYN3_HUMAN	1.68e+00
43	94	3.6	401	1	AROC_MYCTU	AROC_MYCTU	2.28e+00
44	94	3.6	1435	1	DP3A_MYCTU	DP3A_MYCTU	2.28e+00
45	96	3.6	1533	1	PDM_DROME	PDM_DROME	1.23e+00

ALIGNMENTS

RESULT ID	1	STANDARD:	PRT:	345 AA.
AC	P36959;			
DT	01-JUN-1994 (REL. 29, CREATED)			
DT	01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)			
DT	01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)			
DE	GMP REDUCTASE (EC 1.6.6.8) (GUANOSINE 5'-MONOPHOSPHATE OXIDOREDUCTASE).			
GN	GMPR.			
OS	HOMO SAPIENS (HUMAN).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA; OC			
OC	EUTHERIA; PRIMATES.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 89336791.			
RA	KANNO H., HUANG I.Y., KAN Y.W., YOSHIDA A.;			
RL	CELL 58:595-606(1989).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 92098099.			
RA	KONDOH T., KANNO H., CHANG L., YOSHIDA A.;			
RL	HUM. GENET. 88:219-224(1991).			
RN	[3]			
RP	VARIANTS.			
RX	MEDLINE: 92098100.			
RA	KONDOH T., KANNO H., CHANG L., YOSHIDA A.;			
RL	HUM. GENET. 88:225-227(1991).			
CC	-1- FUNCTION: IT FUNCTIONS IN THE CONVERSION OF NUCLEOBASE, NUCLEOSIDE AND NUCLEOTIDE DERIVATIVES OF G TO A NUCLEOTIDE, AND IN MAINTAINING THE INTRACELLULAR BALANCE OF A AND G NUCLEOTIDES.			
CC	-1- CATALYTIC ACTIVITY: NADPH + GUANOSINE 5'-PHOSPHATE -> NADP(+) + INOSINE 5'-PHOSPHATE + NH(3).			
CC	-1- SUBUNIT: HOMOTETRAMER.			
CC	-1- POLYMORPHISM: AT LEAST TWO DIFFERENT ALLELES ARE KNOWN.			
CC	-1- SIMILARITY: TO OTHER EUKARYOTIC AND PROKARYOTIC GMP REDUCTASE AND TO IMP DEHYDROGENASE.			
DR	EMBL: L35304; G528232; -			
DR	EMBL: M27941; G529233; -			
DR	EMBL: M24470; G182867; -			
DR	MIM: 139265; -			
KW	PROSITE: PS00487; IMP_DH_GMP_RED; 1.			
KW	OXIDOREDUCTASE; NADP; POLYMORPHISM.			
FT	BINDING 186 186 IMP (POTENTIAL).			
FT	VARIANT 234 234 A -> T (IN AN ALLELE).			
FT	VARIANT 256 256 F -> I (IN AN ALLELE).			
SO	SEQUENCE 345 AA; 37419 MW; 5DB5E0B CRC32;			

```
D0      1 MPRLEFEKDEKCVLLRPKRSTLRSAEVDLMEYFRRSKTYVVPVVASNMPTGT 60
       |||:::|||||||:::|||||:::|||||:::|||||:::|||||:::
QY     19 MPHLDNDVKRDKFCVLLRPKRSTLSRSEVDLTRSFSSRKQIYSGVPIIAANMTVTGT 78
       |||:::|||||||:::|||||:::|||||:::|||||:::|||||:::
D0      61 FEMAEVLAKFSLETTIRKHVYVDEMKAFFVQVDSNPDMQIGISSISHPDADKLRTVC 120
       |||:::|||||||:::|||||:::|||||:::|||||:::|||||:::
QY     79 FEMAKVLCFESFLAVRHKNHSLVQMOEPAG--NPPCLHLAASSGTSDDPEQEQILI 135
       |||:::|||||||:::|||||:::|||||:::|||||:::|||||:::
D0      121 DMIPLELEYICLDVANGXSEVFEDFIRRRBEOPHTTFAGNVNTGEAVEELITSGADVNY 180
       |||:::|||||||:::|||||:::|||||:::|||||:::|||||:::
QY     136 EAIQVKKTICLDVANGXSEHFVEEVKDVKKRFPQHTIMAGNVATGEAVEELITSGADIK 195
```

```

RP      SEQUENCE FROM N.A.
RC      STRAIN-K12 / W3110.
RX      MEDLINE: 94261430.
RA      FUJITA N., MORI H., YURA T., ISHIHAMA A.;
RL      NUCLEIC ACIDS RES. 22:1637-1639(1994).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN-K12 / MG1655;
RA      BLATTNER F.R., PLUNKETT G. III, MAYHEW G.F., PERNA N.T., GLASNER F.D.;
RL      SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN      [4]
RP      SEQUENCE OF 243-346 FROM N.A.
RC      STRAIN-K12;
RX      MEDLINE: 95047556.
RA      WHITCHURCH C.B., MATYICK J.S.;
RL      GENE 150:9-15(1994).
RN      [5]
RP      SEQUENCE OF 1-12.
RC      STRAIN-K12 / EMG2;
RX      LINK A.J.;
RL      SUBMITTED (OCT-1994) TO THE SWISS-PROT DATA BANK.
CC      -1- FUNCTION: IT FUNCTIONS IN THE CONVERSION OF NUCLEOBASE, NUCLEOSIDES
CC      AND NUCLEOTIDE DERIVATIVES OF G TO A NUCLEOTIDES, AND IN
CC      MAINTAINING THE INTRACELLULAR BALANCE OF A AND G NUCLEOTIDES.
CC      -1- CATALYTIC ACTIVITY: NADPH + GUANOSINE 5'-PHOSPHATE = NADP(+) +
CC      INOSINE 5'-PHOSPHATE + NH(3).
CC      -1- SUBUNIT: HOMOTETRAMER.
CC      -1- SIMILARITY: TO OTHER EUKARYOTIC AND PROKARYOTIC GMP REDUCTASE
CC      AND TO IMP DEHYDROGENASE.
DR      EMBL: X07917; G42616; -.
DR      EMBL: D26562; G43772; -.
DR      EMBL: AE000119; G1786293; -.
DR      EMBL: L28105; G456045; -.
DR      PIR: S01671; S01671.
DR      PIR: S45182; S45182.
DR      ECO2DBASE: H037.4; 6TH EDITION.
DR      ECOGENE: EG10422; GUAC.
DR      PROSITE: PS00487; IMP_DH_GMP_RED; 1.
KW      OXIDOREDUCTASE, NADP.
FT      BINDING      186      IMP (POTENTIAL).
FT      NP_BIND      216      238      NADP (POTENTIAL).

```

FT	CONFLICT	233	234	AR > GCG (IN REF. 3).
SO	SEQUENCE	346 AA;	37440 MW;	A14B3D9B CRC32;
	Query Match		64.9%;	Score 1709; DB 1; Length 346;
	Best Local Similarity	68.4%;	Pred. No. 0.00e+00;	
	Matches	236; Conservative	52; Mismatches	55; Indels 2; Gaps 2
Db	2	RIEEDLKIGFQVLLIRPRKSTLKSRSVDLELROFTFKHSGSGWSGVPILANMDVTGTF	61	
Oy	21	HIDNDVKKDFEDVLLRPRKSTLKSRSDEVDLTRSFSSFRNSKQTSYGVPIIANMDVTGTFE	80	
Db	62	MASLASDILTAAVKKHVSPEMQAFINSSADVLLKHMVSTGSDADFETKRIIDLNP	121	
Oy	81	MAKVLCKSLTAAVKKHSLVQWQEFACQNP-DCLIEHLAASSGSSDFEQLILEAIP	139	
Db	122	ALNFVCLDVANGYSEHFQVFAKKAREAPPTKICAGNVYTGEMCEELLTSGADIVKYG	181	
Oy	140	QVKYICLDVANGYSHFVEFVKDVRKRRPQITMAAGNVYTGEMCEELLTSGADIVKYG	199	
Db	182	PGSVCTTRVKTGVGVPOLSAVIECADAAHGLGMIVSDGCGTTPGVAKAF-ARADFVML	240	
Oy	200	PGSVCTTRKKTGVGVPOLSAVIECADAAHGLGHILISDGCSCPGVAKAFAGADFVML	259	
Db	241	GGMLAGHEESGRIVEENGKFMFLFGMSSSAKKNHVGVAEYRAAEGRVKYKPLAGPV	300	
Oy	260	GGMLAGHSESGELLERDKGRYKFLFGKSSSEMAKKYAGVAEYRAAEGRVKYKPLAGPV	319	
Db	301	ENTARLDIGLRSACTYVAGASRLKELTARTFTFIRQCEIRFIINN 345		
Oy	320	ENTIRLDIGLIRSTCTYVAGAKKELSRITFTFIRYQVANNIFSE 364		
RESULT	4	STANDARD:	PRT:	513 AA.
ID	IMDH_BACSU			
AC	P21879;			
DT	01-MAY-1991 (REL. 18, CREATED)			
DT	01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)			
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)			
DE	INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE (EC 1.1.1.205) (IMP			
DE	DEHYDROGENASE) (IMPDH) (IMPD).			
GN	GUAB OR GNAB.			
OS	BACILLUS SUBTILIS.			
OC	PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-168;			
RX	MEDLINE; 96051385.			
RA	KANZAKI N., MIYAGAWA K.I.;			
RL	NUCLEIC ACIDS RES. 18:6710-6710(1990).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-168;			
RX	MEDLINE; 96051385.			
RA	OGASAWARA N., NAKAI S., YOSHIKAWA H.;			
RL	DNA RES. 1:1-14(1994).			
CC	-1- CATALYTIC ACTIVITY: INOSINE 5'-PHOSPHATE + NAD(+) + H(2)O =			
CC	XANTHOSINE 5'-PHOSPHATE + NADH.			
CC	-1- PATHWAY: FIRST REACTION UNIOUE TO GMP BIOSYNTHESIS.			
CC	-1- SIMILARITY: TO OTHER EUKARYOTIC AND PROKARYOTIC IMPDH AND TO			
CC	GMP REDUCTASE.			
CC	-1- SIMILARITY: CONTAINS 2 CBS DOMAINS.			
DR	EMBL; X55669; G39959; -.			
DR	EMBL; D26185; G46739; -.			
DR	PIR; S12623; DEBSMP.			
DR	SUBTILIST; B610073; GUAB.			
DR	PROSITE; PS00487; IMP DH GMP RED. 1.			
KW	OXIDOREDUCTASE; NAD; GMP BIOSYNTHESIS; PURINE BIOSYNTHESIS; REPEAT;			
KB	CBS DOMAIN.			
FT	DOMAIN	92	147	CBS 1.
FT	DOMAIN	155	209	CBS 2.
FT	BINDING	308	308	IMP (POTENTIAL).
FT	CONFLICT	28	28	H -> R (IN REF. 2).
FT	CONFLICT	480	513	YHRRALDGLGSGHOKTGTVYDECCSGGFSSD ->
FT				KESNTYIS (IN REF. 2).

S0	SEQUENCE	513 AA;	55725 MW;	9816609C CRC32;
	Query Match	20.9%;	Score 550;	DB 1; Length 513;
	Best Local Similarity	38.5%;	Pred. No. 2,61e-98;	
	Matches	99; Conservative	62; Mismatches	89; Indels 7; Gaps 5;
Db	209 EFPNSSKDINHRLVGAAGVTG-DIMTRKAKLEAVANDYVIDTAGHSQGLNTVTKI	267		
Oy	105 EFAGNPDLCEHLAASSGTSGSDEPLEQLIDEAL-PQVKYICLDLVANGYSBHFEVFKDV	163		
Db	268 RETPELNTINGNVAATAEATRALLEAGADVVYKVGISGSICTTRRVAGVGPQTIAIDYC	327		
Oy	164 KRPRPHITMGNVNVTEMEVELLSGDIIKXIGIGSVCTRKKKGVGYPOLSAVMEC	223		
Db	328 ATERRKHGTITADGGIKFSGDITKALAAGHAWMLSLAGTSESPGEIEITYOGRREKV	387		
Oy	224 ADAHHGLKGHIISDGCSCPDVAKAKACAGADFPMLGMLAGHSSESGELLERDKKYKL	263		
Oy	284 FYGMSSEMAAMKRYAG--V-AEYR-AASEKTVEVPKGVHEHTIRDLIGISTSTCYVG	338		
Db	448 SKDRALEAEQAFTMT	464		
Oy	339 AAKLELSRTTFTT	355		
RESULT	5 STANDARD; PRT; 481 AA.			
AC	P56088;			
DT	01-NOV-1997 (REL. 35, CREATED)			
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)			
DE	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)			
DE	INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE (EC 1.1.1.205) (IMP			
DE	DEHYDROGENASE) (IMPDH) (IMPD).			
GN	GMB OR HP0829.			
OS	HELICOBACTER PYLORI (CAMPYLOBACTER PYLORI).			
OC	PROKARYOTA; GRACILICUTES; SCOTOBACTERIA;			
RN	AEROBIC, MOTILE, HELICAL AND/OR VIBRIOID.			
[1]				
RP	SEQUENCE FROM N.A.			
RC	STRAIN-26695;			
RX	MEDLINE; 97394467.			
RA	TOMB J.-F., WHITE O., KERLAUGE A.R., CLAYTON R.A., SUTTON G.G.,			
RA	FELIXSHANN R.D., KETCHUM K.A., KLEK H.-D., GILL S., DOUGHERTY B.A.,			
RA	NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,			
RA	LOTTUS B., RICHARDSON D., DOSON R., KHALK H.G., GLOCK A.,			
RA	MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKET E.K.,			
RA	BENG D.E., GOCAVNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,			
RA	COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATNEY L., WALLIN E.,			
RA	HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,			
RA	VENTER J.C.;			
RU	NATURE 388:539-547(1997).			
CC	-1- CATALYTIC ACTIVITY: INOSINE 5'-PHOSPHATE + NAD(+) + H(2)O -			
CC	XANTHOSINE 5'-PHOSPHATE + NADH.			
CC	-1- PATHWAY: FIRST REACTION UNICUE TO GMP BIOSYNTHESIS.			
CC	-1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).			
CC	-1- SIMILARITY: TO OTHER EUKARYOTIC AND PROKARYOTIC IMPDH AND TO			
CC	GMP REDUCTASE.			
DR	-1- SIMILARITY: CONTAINS 2 CBS DOMAINS.			
DR	PROSITE; PS00487; IMP_DH_GMP_RED; 1.			
FT	TIGR; HP0829; .			
FT	DOMAIN 89 142 CBS 1.			
FT	DOMAIN 150 204 CBS 2.			
FT	BINDING 300 300 IMP (POTENTIAL).			
S0	SEQUENCE 481 AA; 51802 MW; 8C13C596 CRC32;			
	Query Match	20.2%;	Score 532;	DB 1; Length 481;
	Best Local Similarity	38.8%;	Pred. No. 5.40e-94;	
	Matches	100; Conservative	65; Mismatches	83; Indels 10; Gaps 7;
Db	204 EYPANNDPERLRVGAIVGQIDRAE-WL-YKAGVDALVDSAHGSHNIHTLTETIK	261		
[1]				

QY 105 EFAGQNPDCLEHLAASSGTSDFEQLEQLEALPOVKYICLDVANGXSEHFEVFKDVR 164
Db 262 KSLVVDVIV-GNVTYKTSNTSLISAGADATVIGPGSICITRIVAGVGMQVAIDNCV 320
QY 165 KRFQHTIMGNVNTGEMVELLISGADIKVIGPGSVCTTRKKTGTGVQQLAVNECA 224
Db 321 EVASKFPIVADGIRSGVAKALAGASSVMIGSLLAGTESPGDFMIVGGRKSY 380
QY 225 DAHAGLGHITSDGSCSPGVAKAFGAGADFVMLGMLAGHSSGGEILIRDKKTKLF 284
Db 381 RGMISGAMTKGSDRYFOEQEASKEVPEGIEGRVPRYRGVSDMIFOLYGVASSMGYQ 440
QY 285 YGMSSEMAK-K---KY-AGVAEYR-ASEGKTEVPEFGDVHTIRDLIGINSTCTYV 337
Db 441 GAKNILELYQNAEFVEIT 458
QY 338 GAANKELSRRTTFRVT 355

RESULT 6
ID IMDH-STREY STANDARD: PRT: 493 AA.
AC P50099;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE (EC 1.1.1.205) (IMP
DE DEHYDROGENASE) (IMPDH) (IMPD).
GN GUA.
OS STREPTOCOCCUS PYOGENES.
OC PROKARYOTA; FIRMICUTES; COCCI; STREPTOCOCCACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 96084952.
RA ASHBAUGH C.D., WESSELS M.R.;
RL GENE 165:57-60(1995).
CC -1- CATALYTIC ACTIVITY: INOSINE 5'-PHOSPHATE + NAD(+) + H(2)O =
XANTHOSINE 5'-PHOSPHATE + NADH.
CC -1- PATHWAY: FIRST REACTION UNIQUE TO GMP BIOSYNTHESIS.
CC -1- SIMILARITY: TO OTHER EUKARYOTIC AND PROKARYOTIC IMPDH AND TO
GMP REDUCTASE.
CC -1- SIMILARITY: CONTAINS 2 CBS DOMAINS.
DR EMBL: U26056; G924848; .
DR PROSITE: P500487; IMP.DH.GMP.RED.1.
KM OXIDOREDUCTASE; NAD; GMP BIOSYNTHESIS; PURINE BIOSYNTHESIS; REPEAT;
KM CBS DOMAIN.
FT BINDING 310 IMP (POTENTIAL).
SQ SEQUENCE 493 AA; 52807 MW; BD2AF453 CRC32;

Query Match 20.1%; Score 529; DB 1; Length 493;
Best Local Similarity 40.1%; Pred. No. 2.82e-93;
Matches 99; Conservative 58; Mismatches 80; Indels 10; Gaps 5;

Db 225 VAAAVGVTSDTEFERAEALFEAGADA--IVIDTAGHSGAGVARKIAIRHPNFTLLIGN 282
QY 117 LAASSGSSDFEQLEQLEALPOVKYICLDVANGXSEHFEVFKDVRPHTIAGN 176
Db 283 IATREGARALYDAQVDYKVGISGICITRIVAGVGPVDTAIIYDAAVAREGKTIIA 342
QY 177 VVTGEMVEELLISGADIKVIGPGSVCTTRKKTGTGVQQLAVNECADAHAGKHTIS 236
Db 343 DGGIKYSGDIYKALAAAGNAYMLGSMFAGTDEAPGETIYGRKPKYRGMSIAAMKG 402
QY 237 DGGSCSPDVAKAFGAGADFVMLGMLAGHSSGGEILIRGKKTKLYGMSSEMAK- 295
Db 403 SSDRYFOGVSVEANKLYPEGIEGRVAYKGAASDIYFQMLGIRSGMGVYAGDIQELHEN 462
QY 296 ---Y-AGVAEYR-ASEGKTEVPEFGDVHTIRDLIGINSTCTYVGAANKELSR 348
Db 463 AQVEMMS 469
QY 349 TTFIRVT 355

RESULT 7
ID IMDH-BORBU STANDARD: PRT: 404 AA.
AC P49058;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE (EC 1.1.1.205) (IMP
DE DEHYDROGENASE) (IMPDH) (IMPD) (FRAGMENT).
GN GUA.
OS BORRELIA BURGDORFERI (LYME DISEASE SPIROCHETE).
OG PLASMID 26 KB CIRCULAR.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; SPIROCHETES; SPIROCHAETALES;
OC SPIROCHAETACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-B31;
RX MEDLINE: 95050198.
RA MARGOLIS N., HOGAN D., TILLY K., ROSA P.;
RL J. BACTERIOL. 176:6427-6432(1994).
CC -1- CATALYTIC ACTIVITY: INOSINE 5'-PHOSPHATE + NAD(+) + H(2)O =
XANTHOSINE 5'-PHOSPHATE + NADH.
CC -1- PATHWAY: FIRST REACTION UNIQUE TO GMP BIOSYNTHESIS.
CC -1- SIMILARITY: TO OTHER EUKARYOTIC AND PROKARYOTIC IMPDH AND TO
GMP REDUCTASE.
CC -1- SIMILARITY: CONTAINS 2 CBS DOMAINS.
DR EMBL: U13372; G532792; .
DR PROSITE: P500487; IMP.DH.GMP.RED.1.
KM OXIDOREDUCTASE; NAD; GMP BIOSYNTHESIS; PURINE BIOSYNTHESIS; REPEAT;
KM CBS DOMAIN; PLASMID.
FT BINDING 229 IMP (POTENTIAL).
FT NON TER 404
SQ SEQUENCE 404 AA; 43767 MW; 85988E44 CRC32;

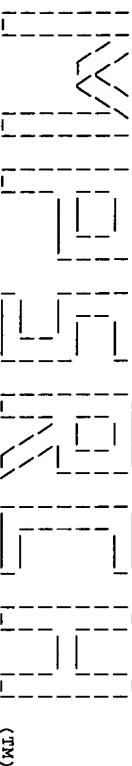
Query Match 18.9%; Score 497; DB 1; Length 404;
Best Local Similarity 36.3%; Pred. No. 1.21e-85;
Matches 95; Conservative 71; Mismatches 86; Indels 10; Gaps 6;

Db 129 EDFNACKD-LNKLRTGAVSDIDIEVEEYKRAHVILYDASHGSTRIELIK 187
QY 104 QEFAGQNPDCLEHLAASSGTSDFEQLEQLEALPOVKYICLDVANGXSEHFEVFKD 162
Db 188 IKTKYPMILDIAGNIVTKEALDLISVADCLKVGISGICITRIVAGVGPVDTAICD 247
QY 163 VKRFPHTIAGNVNTGEMVEELLISGADIKVIGPGSVCTTRKKTGTGVQQLAVNE 222
Db 248 VYECNNNTNICIADGIRSGDVYKAIAGADSVMIIGNFAGTKESPSEIYNGKFK 307
QY 223 CADAHGLKGHITSDGSCSPGVAKAFGAGADFVMLGMLAGHSSGGEILIRDKKTK 282
Db 308 SYVMSGISAMKRSKRRYOLENNEFKKLVPEGIEGRVAYKGAASDIYFQMLGIRSG 367
QY 283 LFYMSSEMAK--K--Y-AGVAEYR-ASEGKTEVPEFGDVHTIRDLIGIRSTC 334
Db 368 GYLGAATISDKINSKFKVISH 389
QY 335 TYGAANKELSRRTTFRVTQ 356

RESULT 8
ID IMDH-METJA STANDARD: PRT: 496 AA.
AC Q59011;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE (EC 1.1.1.205) (IMP
DE DEHYDROGENASE) (IMPDH) (IMPD).
GN GUA OR M1616.
OS METHANOCOCCUS JANNASCHII.
OC ARCHAEABACTERIA; EUBRYARCHAEOTA; METHANOCOCCALES; METHANOCOCCACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 96337999.
RA BUTT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,

Db 397 GSSDRYFQSDNADKLVPEGIEGRVAYKRLKEITHQMGRLS-CMGLTGCCTIDELRT 455
DE INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE (EC 1.1.1.205) (IMP
DE DEHYDROGENASE) (IMPDH) (IMPD).
GN GUAB OR H10221.
OS HAEMOPHILUS INFLUENZAE.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC PASTEURRELLACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-EB 104.
RA ANDEREGG U., SCHNUCK W.H., ASPERGER O., KLEBER H.P.;
RL SUBMITTED (JUN-1992) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- CATALYTIC ACTIVITY: INOSINE 5'-PHOSPHATE + NAD(+) + H(2)O =
CC XANTHOSINE 5'-PHOSPHATE + NADH.
CC -1- SIMILARITY: TO OTHER EUKARYOTIC AND PROKARYOTIC IMPDH AND TO
CC -1- SIMILARITY: FIRST REACTION UNIQUE TO GMP BIOSYNTHESIS.
CC GMP REDUCTASE.
CC -1- SIMILARITY: TO OTHER EUKARYOTIC AND PROKARYOTIC IMPDH AND TO
CC GMP REDUCTASE.
CC -1- SIMILARITY: CONTAINS 2 CBS DOMAINS.
DR EMBL; X66859; G38720; -.
DR PROSITE; PS00487; IMP_DH_GMP_RED; 1.
KM OXIDOREDUCTASE; NAD; GMP BIOSYNTHESIS; PURINE BIOSYNTHESIS; REPEAT;
KW CBS DOMAIN.
FT DOMAIN 90 142 CBS 1.
FT BINDING 150 205 CBS 2.
FT SEQUENCE 304 304 IMP (POTENTIAL).
SQ SEQUENCE 488 AA; 51530 MW; 1D4AA72F CRC32;
Query Match 18.0%; Score 474; DB 1; Length 488;
Best Local Similarity 38.9%; Pred. No. 3,44e-80;
Matches 98; Conservative 69; Mismatches 68; Indels 17; Gaps 11;
Db 218 RVGAAGTGTADTPSRVETALVAGVDV--IVVDTAHGHAGVIERVYKONFPQOVITGG 275
QY 116 HLAASSTGSSDFEQLDLEALPQVYKICLDVANGYSEHVEVEKDYKRRFPQHTIMAG 175
Db 276 NIATGDAALLDAGADAVKVGIGPGSICITRTIVAGIGMPOISAI-DSV-AS-ALKDQIP 332
QY 176 NVVGEWEEELILSGADILIKYIGIGSVCTIRKKTGCVGQPOLSAVMECADAHGLKGIH- 234
Db 333 LIAOGGIFSGDMKAKIAGASTIVGSLAGTEAPGEVEFPGQRYKAKRKGSLGAM 392
QY 235 -ISDGGCCPGDVAFAFGADFWLGLAGHSGSEGLIERDQKKYKLLPYGMSSEVAM 293
Db 393 AGRTGSADRYFQSDNADKLVPEGIEGRVAYKRLKEITHQMGRLS-CMGLTGCCTIDELR 452
QY 294 KRYAGVAVET-RAS-EG--K--T--VE--VFPKGVETHTINDIGISTCTTYGAANKLK 343
Db 453 DLRONAKFVKIT 464
QY 344 ELSRRTFIRVT 355
RESULT 11
ID IMDH_HAEIN STANDARD; PRT; 488 AA.
AC P44334;
DT 01-NOV-1995 (REL. 32, CREATED)
DE INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE (EC 1.1.1.205) (IMP
DE DEHYDROGENASE) (IMPDH) (IMPD).
GN GUAB.
OS ACINETOBACTER CALCOACETICUS.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;
OC NEISSERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-EB 104.
RA ANDEREGG U., SCHNUCK W.H., ASPERGER O., KLEBER H.P.;
RL SUBMITTED (JUN-1992) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- CATALYTIC ACTIVITY: INOSINE 5'-PHOSPHATE + NAD(+) + H(2)O =
CC XANTHOSINE 5'-PHOSPHATE + NADH.
CC -1- SIMILARITY: TO OTHER EUKARYOTIC AND PROKARYOTIC IMPDH AND TO
CC -1- SIMILARITY: FIRST REACTION UNIQUE TO GMP BIOSYNTHESIS.
CC GMP REDUCTASE.
CC -1- SIMILARITY: TO OTHER EUKARYOTIC AND PROKARYOTIC IMPDH AND TO
CC GMP REDUCTASE.
CC -1- SIMILARITY: CONTAINS 2 CBS DOMAINS.
DR EMBL; X66859; G38720; -.
DR PROSITE; PS00487; IMP_DH_GMP_RED; 1.
KM OXIDOREDUCTASE; NAD; GMP BIOSYNTHESIS; PURINE BIOSYNTHESIS; REPEAT;
KW CBS DOMAIN.
FT DOMAIN 90 142 CBS 1.
FT BINDING 150 205 CBS 2.
FT SEQUENCE 304 304 IMP (POTENTIAL).
SQ SEQUENCE 488 AA; 51530 MW; 1D4AA72F CRC32;

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE (EC 1.1.1.205) (IMP
DE DEHYDROGENASE) (IMPDH) (IMPD).
GN GUAB OR H10221.
OS HAEMOPHILUS INFLUENZAE.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC PASTEURRELLACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20;
RX MEDLINE; 95350630.
RA FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
RA KERLAVAGE A.R., BULL C.J., TOMB J.F., DOUGHERTY B.A., MERRICK J.M.,
RA MCKENNEY K., STUTON G., FITZHUGH W., FIELDS C.A., GGCAYNE J.D.,
RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODER A., KELLEY J.M.,
RA WETTERMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOW E., COTTON M.D.,
RA UETTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
RA GNEHM C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
RA VENTER J.C.;
RL SCIENCE 269:496-512(1995).
CC -1- CATALYTIC ACTIVITY: INOSINE 5'-PHOSPHATE + NAD(+) + H(2)O =
CC XANTHOSINE 5'-PHOSPHATE + NADH.
CC -1- PATHWAY: FIRST REACTION UNIQUE TO GMP BIOSYNTHESIS.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SIMILARITY: TO OTHER EUKARYOTIC AND PROKARYOTIC IMPDH AND TO
CC GMP REDUCTASE.
CC -1- SIMILARITY: CONTAINS 2 CBS DOMAINS.
DR EMBL; U32708; G1573185; -.
DR PROSITE; PS00487; IMP_DH_GMP_RED; 1.
DR TIGR; H10221; -.
KM OXIDOREDUCTASE; NAD; GMP BIOSYNTHESIS; PURINE BIOSYNTHESIS; REPEAT;
KW CBS DOMAIN.
FT DOMAIN 91 144 CBS 1.
FT BINDING 152 207 CBS 2.
FT SEQUENCE 306 306 IMP (POTENTIAL).
SQ SEQUENCE 488 AA; 51981 MW; 79CFB63E CRC32;
Query Match 18.0%; Score 474; DB 1; Length 488;
Best Local Similarity 36.9%; Pred. No. 3,44e-80;
Matches 92; Conservative 68; Mismatches 76; Indels 13; Gaps 10;
Db 220 RVGAAGAGAGNERIDALVKAGVDV-LI-IDSSHSGSEGLDVRRTTRAKRPLPIVAG 277
QY 116 HLAASSTGSSDFEQLDLEALPQVYKICLDVANGYSEHVEVEKDYKRRFPQHTIMAG 175
Db 278 NVATAGGAIALADAGASAVKVGIGPGSICITRTIVTGVGPQITADAAALKD-RGIPV 336
QY 176 NVVGEWEEELILSGADILIKYIGIGSVCTIRKKTGCVGQPOLSAVMECADAHGLKGIH- 234
Db 337 IADGIRFSGDIKAKIAGASCVVWGMFAGTEAPGEIELLYOGRAKSYRKGSLGAMA 396
QY 235 ISDGGCCPGDVAFAFGADFWLGLAGHSGSEGLIERDQKKYKLLPYGMS--EMA 292
Db 397 KGSDDRYFQSDNADKLVPEGIEGRVAYKRLKEITHQMGRLS-CMGLTGCCTIDELR 455
QY 293 M--KKI-AGGVAYET-RASEKTYVEVPKGVDEHTIDIGISTCTTYGAANKLK 346
Db 456 TKAEFVRIS 464
QY 347 RRTTFIRVT 355
RESULT 12
ID IMDH_PYRFU STANDARD; PRT; 485 AA.
AC P42851;
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE (EC 1.1.1.205) (IMP
DE DEHYDROGENASE) (IMPDH) (IMPD).
GN GUAB.
OS PYROCOCCUS FURIOSUS.



Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Mpsrch_p protein - protein database search, using Smith-Waterman algorithm
 Run on: Mon Jun 15 16:00:50 1998; Maspar time 19.23 Seconds
 Tabular output not generated. 801.463 Million cell updates/sec

Title: >US-08-774-169-1
 Description: (1-366) from US08774169.pep
 Perfect Score: 2632
 Sequence: 1 MTCCPAPARFATPRLSAMP.....RRTRFRVYQVNPISFSEAC 366

Scoring table:
 Gap 11
 PAM 150

Searched: 140542 seqs, 42109429 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: sptrembl5
 1:sp_fungi 2:sp_human 3:sp_invertebrate 4:sp_mammal
 5:sp_mhc 6:sp_organelle 7:sp_phase 8:sp_plant
 9:sp_bacteria 10:sp_rodent 11:sp_virus 12:sp_invertebrate
 13:sp_unclassified

Statistics: Mean 47.716; Variance 81.544; scale 0.585

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	1852	70.4	358	3	016294	F32D1.5 PROTEIN.	0.00e+00
2	529	20.1	326	9	005269	HYPOTHETICAL 35.8 KD P	7.63e-91
3	466	17.7	484	9	026245	INOSINE-5'-MONOPHOSPHA	2.61e-76
4	454	17.2	478	9	032912	INOSINE-5'-MONOPHOSPHA	1.45e-73
5	432	16.4	327	9	025525	GMP REDUCTASE (GPOAC).	1.49e-68
6	365	13.9	524	1	014344	HYPOTHETICAL 57.0 KD P	1.74e-53
7	201	7.6	387	9	P73853	IMP DEHYDROGENASE SUBU	9.89e-19
8	111	4.2	487	9	033941	ERYTH.	1.09e-02
9	104	4.0	175	2	016686	AMP DEAMINASE (EC 3.5.	1.19e-01
10	104	4.0	256	2	016687	AMP DEAMINASE (EC 3.5.	1.19e-01
11	105	4.0	357	9	051741	COFACTOR MODIFYING PRO	8.50e-02
12	104	4.0	605	2	016688	AMP DEAMINASE (EC 3.5.	1.19e-01
13	104	4.0	753	2	016729	AMP DEAMINASE ISOFORM	1.19e-01
14	103	3.9	342	1	006148	CHROMOSOME XII COSMID	1.66e-01
15	102	3.9	511	9	029309	GLUTAMATE SYNTHASE (GL	2.30e-01
16	102	3.9	622	9	026208	GLUTAMATE SYNTHASE (NA	2.30e-01
17	99	3.8	295	3	001390	SYNTAXIN 1 HOMOLOG.	6.10e-01
18	100	3.8	307	3	024253	SIMILAR S. CEREVISIAE	4.42e-01
19	97	3.7	126	3	024125	CYCLOCHROME P450 (FRAGM	1.15e+00
20	97	3.7	243	9	031677	KVIL PROTEIN.	1.15e+00

21	97	3.7	661	9	033900	HYPOTHETICAL 72.6 KD P	1.15e+00
22	98	3.7	722	10 <td>035826</td> <td>UDP-N-ACETYLGLUCOSAMIN</td> <td>8.40e-01</td>	035826	UDP-N-ACETYLGLUCOSAMIN	8.40e-01
23	97	3.7	1083	1	006108	CHROMOSOME XVI COSMID	1.15e+00
24	98	3.7	1741	3	019206	CODED FOR BY C. ELEGAN	8.40e-01
25	95	3.6	248	3	015562	GLUTAMYL-TRNA SYNTHETA	2.16e+00
26	94	3.6	282	3	002335	MO2D9.2.	2.94e+00
27	94	3.6	345	11	067663	GENOMIC RNA.	2.94e+00
28	96	3.6	433	9	005093	PUTATIVE SUGAR-BINDING	1.58e+00
29	94	3.6	2890	9	025806	DNA-DIRECTED RNA POLYM	2.94e+00
30	91	3.5	256	9	P94848	29 KDA PROTEIN.	7.31e+00
31	92	3.5	260	8	041450	CYCLOCHROME C1 (FRAGMEN	5.41e+00
32	91	3.5	274	11	089048	(STRAIN 01/85-PN-B1) P	7.31e+00
33	92	3.5	289	8	004240	RNA-OR SSDNA-BINDING	5.41e+00
34	93	3.5	317	6	019892	YCF7 PROTEIN.	4.00e+00
35	92	3.5	320	8	041207	CYCLOCHROME C1 (FRAGMEN	5.41e+00
36	92	3.5	366	9	028226	2-KETOISOLATE RATE FERR	5.41e+00
37	91	3.5	411	9	027121	3-PHOSPHOGLYCERATE KIN	7.31e+00
38	92	3.5	422	8	040543	PROTEIN-SERINE/THREONIN	5.41e+00
39	92	3.5	433	9	P94528	HYPOTHETICAL 48.6 KD P	5.41e+00
40	92	3.5	519	3	009592	HYPOTHETICAL 57.0 KD P	5.41e+00
41	92	3.5	520	8	042600	CYCLOCHROME P450 (FERUL	5.41e+00
42	91	3.5	532	10	060849	CD98 HEAVY CHAIN.	7.31e+00
43	93	3.5	595	9	029490	TRANSLATION INITIATION	4.00e+00
44	93	3.5	928	9	032556	SENSORY KINASE.	4.00e+00
45	91	3.5	1379	3	P91824	T22A3.8.	7.31e+00

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	358 AA.
ID	016294;			
AC	016294;			
DT	01-JAN-1998 (TREMBLREL. 05, CREATED)			
DT	01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)			
DT	01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)			
DE	F32D1.5 PROTEIN.			
GN	F32D1.5.			
OS	CAENORHABDITIS ELEGANS.			
OC	ERKAROTIA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BRISTOL N2;			
RX	MEDLINE; 94150718.			
RA	WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,			
RA	BONFIELD J., BURTON J., CONNELL M., COPESEY T., COOPER J., COULSON A.,			
RA	CAXTON M., DEAR S., DU Z., DURBIN R., FAYELLO A., FULFON L.,			
RA	GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,			
RA	JONES W., KERSHAW J., KIRSTEN J., LAISTER J., LATREILLE P.,			
RA	LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,			
RA	PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,			
RA	SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SUTSTON J.,			
RA	THIERRY-MIEG J., THOMAS K., VAUDIN M., VANGHAN K., WATERSTON R.,			
RA	WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLIDAN P.,			
RL	NATURE 368:32-38(1994).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BRISTOL N2;			
RA	BECKER M., BRADSHAW H., KRAMER J.;			
RL	SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.			
DR	EMBL; AF016427, G2291231. -			
DR	PROSITE; PS00487; IMP_DH_GMP_RED; 1.			
SO	SEQUENCE 358 AA; 38806 MW; 1A112102 CRC32;			
Query Match	70.4%; Score 1852; DB 3; Length 358;			
Best Local Similarity	70.4%; Pred. No. 0.00e+00;			
Matches	243; Conservative 60; Mismatches 41; Indels 1; Gaps 1;			
Db	1 MPRIENPKLDEKDVLLPRKSTLKSRAVDLEDRVFRNSKATYTGVPVVASNMOTVGT 60			

```

* 19 PHRIDNDVKLDFKVLARPKRSTLKSREVDLTRSFSSRNKQITSGPPIIANNNDTGT 78
Db 61 FEMAAALNNHKIFTTIHKHYVDENKAPASASPDTENNLAISGISNDWTKLTVTITE 120
Qy 79 FEMAKVLCFKFSLFTAVHSHYLVQWQEFAG-QNPDLCEHLAASSGTSGSDEQLEQILDEA 137
Db 121 LPQKTYICLDVANGSSESEFIRVRREAPYKPHITMAENVYTGEMVELLISGADIVAVG 180
Qy 138 IPQKTYICLDVANGSSESEFIRVRREAPYKPHITMAENVYTGEMVELLISGADIVAVG 197
Db 181 IGPSSVCTTRKAGVGYQPOLSAVLECAADAAGLNGHVMASDGCSPGVAAKAFGADFEV 240
Qy 198 IGPSSVCTTRKAGVGYQPOLSAVLECAADAAGLNGHVMASDGCSPGVAAKAFGADFEV 257
Db 241 MIGGLFAGHDOSGGDLIHNGKFKFLFYGMSSDPAKMKHGSVAEYRASEGKVTYTPYRG 300
Qy 258 MIGGLFAGHDOSGGDLIHNGKFKFLFYGMSSDPAKMKHGSVAEYRASEGKVTYTPYRG 317
Db 301 DVNGTVODILGIRACVYTGAKHKLAKRATFIRVYQQTNDWY 345
Qy 318 DVNTRIDILGIRSTCTYVGAAKLKLKSLRRTTFRVYQVNPFR 362

RESULT 2
ID 005269 PRELIMINARY: PRT: 326 AA.
AC 005269;
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 35.8 KD PROTEIN.
GN YUWD.
OS BACILLUS SUBTILIS.
OC PROKARYOTA: FIRMICTES: ENDOSPORE-FORMING RODS AND COCCI: BACILLACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RN [2]
RP SEQUENCE FROM N.A.
RC KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONT G.,
RA AZEVEDO V., BERTERO M.G., BESIETTES P., BOLOTIN A., BORCHERT S.,
RA BOURIET R., BOUSIER L., BRANS A., BRAUN M., BRIGELL S.C., BRON S.,
RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,
RA DERIZOT F., DEVINE K.M., DUSTERHOFF A., EHRLICH S.D., EMMERSON P.T.,
RA ENTIAN K.D., ERINGTON J., FABRET C., FERRARI E., FOULGER D.,
RA FERTZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALIERON N.,
RA GHM S.Y., GLASER P., GOFEDAU A., GOLIGHTLY E.J., GRANDI G.,
RA GUISEPI G., GUY B.J., HAGA K., HAIBCH J., HARWOOD C.R., HENAUT A.,
RA HILBERT H., HOLAPPEL S., HOSONO S., HURLLO M.F., ITAYA M., JONES L.,
RA JOIS B., KAKAMATA D., KASHARA Y., KLAER-BLANCHARD M., KLEIN C.,
RA KOSAVSHI Y., KOTTER P., KONINGSTEIN G., KROGH S., KUMANO M.,
RA KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
RA MEDINA N., MELLAO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBAC M.,
RA MOONE D., O'REILLY M., OGAWA K., OGAWA K., ODEGA B., PARK S.H.,
RA PARO V., POHL T.M., PORTERELLE D., PORROLIK S., PRESCOTT A.M.,
RA PRESECAN E., PUJIC P., PUENELLE B., RAPPORT G., REY M., REYNOLDS S.,
RA RIGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADIE Y.,
RA SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOPFONE F.,
RA SERIGUCHI J., SKOWSKA A., SEROR S.J., SEROR P., SHIN B.S., SOLDO B.,
RA SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
RA TAKEUCHI M., TANAKASHI A., TANAKA T., TERPSTRA P., TOGNONI A.,
RA TOSATO V., UCHITAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
RA VIARI A., WABUTT R., WEDLER E., WEDLER H., WETZNEGER T.,
RA WITKES P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.,
RL NATURE 390:249-256(1997).
RN [3]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-168;
RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; 293939; E311468; -.
DR EMBL; 299120; E1184292; -.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 326 AA; 35819 MW; 4566014F CRC32;

Query Match
Best Local Similarity 35.3%; Score 529; DB 9; Length 326;
Matches 118; Conservative 79; Mismatches 116; Indels 21; Gaps 17;

Db 5 FDDYDIOLIPAKCVNSRSEC-T-S-V-RIGGTEK-LPVPANMOTIIDKRLAELAE 59
Qy 28 LDFKDVILRPKRSTLKSREVDLTRSFSSRNKQITSGPPIIANNNDTGTFFEMAKVLC 87
Db 60 NGFYVMHR-FEPTRIDFIDKNNARGL-FSSISGVKDEYEFVROLAENLTPEVYTI 117
Qy 88 FSLFTAVHSHYLVQWQEFAG-QNPDLCEHLAASSGTSGSDEQLEQILDEAIPQVKTICL 146
Db 118 DIANGHNAVETMIQHLKHLIPDSFVIAGVNGPREAVRELEMGADATKVGIGPKVCIT 177
Qy 147 DVANGSEHFEVFKDYKRREPQHTIMAGNVYTGEMVELLISGADIIKYGIGPGSVCTT 206
Db 178 KIKGFGTGWQLAALRWAKAA-S-K-PIIADGICITHGDIAKSIKATMVMIGSLFA 234
Qy 207 RKTGTVGP-QLSAVMECAADAAGLNGHVMASDGCSPGVAAKAFGADFEVNLGGLMA 264
Db 235 GHESPQGTIEKDKLKYEGSASEPP-K--G-EK-KNVEGKNNVHAKGSIXKPTLI 287
Qy 265 GHSESGELIRDRCKTKKFLFGMSSEMAKMYAGVAVYRASEKTYEVFPKGVENHTR 324
Db 288 EMEODLOSSISYAGCTKLNAL-RNVDYIVYKNSI 320
Qy 325 DILGIRSTCTYVGAAKLKLKSLRRTTFRVYQV 358

RESULT 3
ID 026245 PRELIMINARY: PRT: 484 AA.
AC 026245;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE.
GN MTH142.
OS METHANOBACTERIUM THERMOAUTOTROPHICUM.
OC ARCHAEABACTERIA: EURYARCHAEOTA; METHANOBACTERIALES;
OC METHANOBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DELTA H;
RA SMITH D.R., DOUCETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DUBOIS J.,
RA ALDREGE T., BASHIRZADEH R., BLAKELY D., COOK R., GILBERT K.,
RA HARRISON D., HOANG L., KEAGLE P., LUMM W., POHNER B., OIU D.,
RA SPRADORA R., VTCARE R., WANG Y., WIERZBOWSKI J., GIBSON R.,
RA JIANG N., CARUSO A., BUSH D., SAHER H., PATWEL J.D., PRABHAKAR S.,
RA MCDONALD S., SHIMER G., GOVAL A., PIETROVSKI S., CHURCH G.M.,
RA DANIELS C.J., MAO J.-I., RICE P., NOLLING J., REEVE J.N.;
RL J. BACTERIOL. 179:7135-7155(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-DELTA H;
RA SMITH D.R.;
RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AE000803; G2621183; -.
SQ SEQUENCE 484 AA; 52206 MW; 3DBCC8C8 CRC32;

Query Match
Best Local Similarity 36.1%; Score 466; DB 9; Length 484;
Matches 97; Conservative 68; Mismatches 88; Indels 16; Gaps 9;

Db 203 ILERKRYPNASRDESGVILRVAAAGTGFDLER-ARALDE-AGADVLAIDSAGHNMNLYKS 260

```

```

OY 100 LVQWQFAGONPDCLEHLAASSTGSDFEQLEILAIPOVKYICLDVANGSEHVEF 159
DB 261 AGAKKEIEADLY-GNATREAEADLIADVDGLKYGSGSCTRIIAGVCPOLTA 319
OY 160 VKDVRKFPQHTIMAGNVYGEWVEELISGADILIKYIGIGSVCTTRKTGVPOLSA 219
DB 320 IAEVADVAEYGVPIADGIRISGDIKAKAIAGADCVMLGNLLAGTYEAPGVVYNGR 379
OY 220 VMECADAAHGLKHIIISDGCSCPGDVAKAFAGADFVMLGMLAGHSESGELIERDGK 279
DB 380 KYKQYRGMGSLGANTGIGAGTDRYFQEPBRHMKTKVPEBGVGVPIKGTSEVLFOL 439
OY 280 KYKLEFYMSSMAKKKAYGVAE-Y-FASEG--R-T-----VE--VPFKGDVEHTIRDI 326
DB 440 IGLRASMGCYCAANTSEMKEKRAIVRT 468
OY 327 LGGIRSTCTYGAAKKELSRRTFIRVT 355

RESULT 4
ID 032912 PRELIMINARY; PRT: 478 AA.
AC 032912:
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE.
GN GUAB.
OS MYCOBACTERIUM LEPRAE.
OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RA SKELTON J., CHURCHER C.M.;
RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA PAKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RA EIGLMEIER K., HONORE N., WOODS S.A., CAUDRON B., COLE S.T.;
RL MOL. MICROBIOL. 7:197-206(1993).
DR EMBL: AL008609; E1168632;
SQ SEQUENCE 478 AA; 50383 MW; B8FC1940 CRC32;

Query Match 17.2%; Score 454; DB 9; Length 478;
Best Local Similarity 42.3%; Pred. No. 1,45e-73;
Matches 104; Conservative 50; Mismatches 79; Indels 13; Gaps 10;

DB 214 RLRTGAAGVINGDVAGRAQSL-AEAGVDVLTVDTAHGVOVTELEIKCVASLNLGVPLVA 272
OY 116 HLAASSGTG--SSDEQLEQLEALIPVKYICLDVANGYSEHFEVDRKRFPQHTIMA 174
DB 273 GNVYSAGSTRLLNAGATYKVGPGAMCTTRMTGVRQFSVAVLECAASAKLNRYV 332
OY 175 GNVYTGVEVELLISGADILIKYIGIGSVCTTRKTGVPOLSAVMECADAAHGLGHI 234
DB 333 WADGVHPRDVALALALAGASNNVIGSWFAGTYPSPGLMDRHDODPYKESYGASRAV 392
OY 235 ISDGCSCPGDVAKAFAGADFVMLGMLAGHSESGELI-ERQCKKTKLFTYGASSEAM 293
DB 393 VARTVADSSDFARRAKALFDEGISTSRMGLDPDHGVEDLDIHITSGVYSTCTYGAASLTA 452
OY 294 --KRYAGVAE-YR-A--SEG-KT--VEV-PFKGDVEHTIRDIIGIRSTCTYGAANKLK 343
DB 453 ELHEKA 458
OY 344 ELSRRT 349

RESULT 5
ID 035525 PRELIMINARY; PRT: 327 AA.
AC 035525:
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)

```

```

DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DT GMP REDUCTASE (GUAC).
GN HP0854.
OS HELICOBACTER PYLORI (CAMPILOBACTER PYLORI).
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA;
OC AEROBIC, MOTILE, HELICAL AND/OR VIBRIOID.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-26695;
RA TOMB, WHITE, KERLAVAGE, CLAYTON, SUTTON, FLEISCHMANN, KETCHUM, KLENK,
RA GILL, DOUGHERTY, NELSON, QUACKENBUSH, ZHOU, KIRKNESS, PETERSON, LOFUS,
RA RICHARDSON, DODSON, KHALAK, GLODER, MCKENNEY, FITZGERALD, LEE, ADAMS,
RA HICKEY, BERG, GOCAYNE, UTTERBACK, PETERSON, KELLEY, COTTON, WEIDMAN,
RA FUJII, BOWMAN, WATTHEY, WALLIN, HAYES, BORODOVSKY, KARP, SMITH,
RA FRASER VENTER.;
RL NATURE 388:539-547(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-26695;
RA TOMB, WHITE, KERLAVAGE, CLAYTON, SUTTON, FLEISCHMANN, KETCHUM, KLENK,
RA GILL, DOUGHERTY, NELSON, QUACKENBUSH, ZHOU, KIRKNESS, PETERSON, LOFUS,
RA RICHARDSON, DODSON, KHALAK, GLODER, MCKENNEY, FITZGERALD, LEE, ADAMS,
RA HICKEY, BERG, GOCAYNE, UTTERBACK, PETERSON, KELLEY, COTTON, WEIDMAN,
RA FUJII, BOWMAN, WATTHEY, WALLIN, HAYES, BORODOVSKY, KARP, SMITH,
RA FRASER VENTER.;
RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AE000596; G2313987;
SQ SEQUENCE 327 AA; 36038 MW; PAD41044 CRC32;

Query Match 16.4%; Score 432; DB 9; Length 327;
Best Local Similarity 32.3%; Pred. No. 1,49e-68;
Matches 108; Conservative 79; Mismatches 126; Indels 21; Gaps 15;

DB 6 FYEDVQILNPKNCTVNSRSCD-TTVI-L--GKHAIF-MPIVANNQTIINESIAEFLAE 60
OY 28 LDFKDVLLRKRSTLKSREVDLTRSFSEFNSKQTSYGVPIIANNQDTGFEMAYLCK 87
DB 61 NGFYIYHNR-FDGAARFPFKMKKR--QWISSTSVGVKKECFVELAKOGIAPYIYI 118
OY 88 FSLPTAVHKHYSLVQWEEFGONPDCLEHLAA--SSGSSDFFQLEILAIPOVKYICL 146
DB 119 DIAHGSNSYIEMIORIKTLPTFVLAVNGVTPREAVRELEENAGDATKYGIGPGKVCIT 178
OY 147 DVANGYSEHVEFVKDVRKFPQHTIMAGNVYGEWVEELISGADILIKYIGIGSVCTT 206
DB 179 KITGFGTGMQALALRWCARKAR--K-PIIADGIRTHGDIVKSIKFGATMYMIGSLFA 235
OY 207 RKTGAGYV--QLSAVMECADAAHGLKHIIISDGCSCPGDVAKAFAGADFVMLGMLA 264
DB 236 GHESSEGTIEENGIAKKEFGSASEF--QK---G-EK-KNIGSKNIWIOHKSLSLDIVL 288
OY 265 GHESSEGTIERGKRYKLEFGSSSEAMKRYAGVAEYRASGKTVEVFPKGDVEHTIR 324
DB 289 EHMDDLOSSISYAGRDLEAI-RKVDYVIKNSI 321
OY 325 DILGIRSTCTYGAANKKELSRRTFIRTOY 358

RESULT 6
ID 014344 PRELIMINARY; PRT: 524 AA.
AC 014344:
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 57.0 KD PROTEIN.
GN SPBC2F12.14C.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972H-7;
RA WOOD V., RAJANDREAM M.A., BARRELL B.G., SKELTON J., CHURCHER C.M.;

```



```

RX TISSUE=CREBBLLUM;
RC MEDLINE: 961033174.
RA DEN BERGH F., SABINA R.L.;
RL BIOCHEM. J. 312:401-410(1995).
CC -1- CATALYTIC ACTIVITY: AMP + H(2)O = IMP + NH(3).
DR EMBL: U16269; G608497; -.
KW HYDROLASE.
FT NON_TER      1      1
FT SEQUENCE     605 AA; 70771 MW; B938C303 CRC32;
SQ
Query Match          4.0%; Score 104; DB 2; Length 605;
Best Local Similarity 50.0%; Pred.No.1,19e-01;
Matches 15; Conservative 7; Mismatches 7; Indels 1; Gaps 1;

Db    47 QISODVKLE-PDILLRAKODEFLKTDSDDL 75
      :| ||||: |:|::|: |::|: |::|:
Oy   21 HIDNDVKLDFKDVLRRPKRSTLKRSSEVDL 50

RESULT 13
ID O16729 PRELIMINARY; PRT; 753 AA.
AC O16729;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE AMF DEAMINASE ISOFORM L SPLICING VARIANT (EC 3.5.4.6) (MYOADENYLATE
DE DEAMINATION) (AMP AMINASE) (ADENYLYLIC ACID DEAMINASE) (FRAGMENT).
GN AMD2.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAEOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA FETAL HEART;
RA BAUSCH-JURKEN M.T., MAHNKE-ZITZELMAN D.K., MORISATO T., SABINA R.L.;
RL SUBMITTED (AUG-1992) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA FETAL HEART;
RA DEN BERGH F., SABINA R.L.;
RL SUBMITTED (AUG-1992) TO EMBL/GENBANK/DBJ DATA BANKS.
CC CC -1- CATALYTIC ACTIVITY: AMP + H(2)O = IMP + NH(3).
DR EMBL: M91029; G644509; -.
KW HYDROLASE.
FT NON_TER      1      1
FT SEQUENCE     753 AA; 87383 MW; 1AE7A7EB CRC32;
SQ
Query Match          4.0%; Score 104; DB 2; Length 753;
Best Local Similarity 50.0%; Pred.No.1,19e-01;
Matches 15; Conservative 7; Mismatches 7; Indels 1; Gaps 1;

Db    40 QISODVKLE-PDILLRAKODEFLKTDSDDL 68
      :| ||||: |:|::|: |::|: |::|:
Oy   21 HIDNDVKLDFKDVLRRPKRSTLKRSSEVDL 50

RESULT 14
ID O06148 PRELIMINARY; PRT; 342 AA.
AC O06148;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE CHROMOSOME XII COSMID 8479.
GN LB479.12.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMTASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C (AB972);
RA JOHNSON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DU Z.,
RA FANELLO A., FULION L., GATTUNG S., GRECO T., KIRSTEN J.,
RA KUOABA T., HALLSWORTH K., HAWKINS J., HILTIER L., JIER M.,
```

```

RA JOHNSON D., JOHNSTON L., LANGSTON Y., LATREILLE P., LE T.,
RA MARDIS E., MENEZES S., MILLER N., NHAN M., PALEY A., PELUSO D.,
RA RIFKEN L., RILES L., RAICH A., TREVASKIS E., VIGNATI D.,
RA WILCOX L., WOHLDMAN P., VAUDIN M., WILSON R., WATERSTON R.,
RL SUBMITTED (NOV-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C (AB972);
RA MILLER N.;
RL SUBMITTED (NOV-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C (AB972);
RA WATERSTON R.;
RL SUBMITTED (NOV-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U1744; G577183; -
SQ SEQUENCE 342 AA; 39115 MW; A3AF6D52 CRC32;

Query Match          3.9%: Score 103; DB 1; Length 342;
Best Local Similarity 23.4%; Pred. No. 1.66e-01;
Matches 22; Conservative 26; Mismatches 40; Indels 6; Gaps 5;

Db 80 IKLKMLDILAPQEIINGVLVSNLY--FSFVSADITVSINSVSHV-TKDMFYMIQS 136
   ||::||| | | | | | | | | | | | | | | | | | | | | | | | | |
QY 26 VKLDFKDVILRKPRSTLKRSEVDLTRSFSPRNKQTYSGVPIIAANDTYGTFEMA-KV 84
   ||::| | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 137 LCMALLKLVNSTQHYVQR-DIINEKOKCLDFL 169
   ||::| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 85 LCKFSL-FTAVHKHYSLVQWQEFAGQNPDCLEHL 117

RESULT 15
ID 029309 PRELIMINARY; PRT: 511 AA.
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE GLUTAMATE SYNTHASE (GLTB).
GN AF0953.
OS ARCHAEoglobus fulgidus.
OC ARCHAEOBACTERIA; EURYARCHAEOTA; ARCHAEoglobales; ARCHAEoglobaceae.
RN [1]
RP SEQUENCE FROM N.A.
RA KLENK H.P., CLAYTON R.A., TOMB J., WHITE O., NELSON K.E.,
RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIDES N.C.,
RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODER A., ZHOU L.,
RA OVERBERK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,
RA COTTON M.D., SPRIGGS T., ARTACH P., KAINE B.P., SYKES S.M.,
RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
RA VENTER J.C.;
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA KLENK H.P., CLAYTON R.A., TOMB J., WHITE O., NELSON K.E.,
RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIDES N.C.,
RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODER A., ZHOU L.,
RA OVERBERK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,
RA COTTON M.D., SPRIGGS T., ARTACH P., KAINE B.P., SYKES S.M.,
RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
RA VENTER J.C.;
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AE001038; G2649642; -
DR PROSITE; PS00198; 4FE4S_FERRDOXIN; 2.
KW IRON-SULFUR.
SQ SEQUENCE 511 AA; 56203 MW; F8C54EF5 CRC32;

```

```

Query Match          3.9%: Score 102; DB 9; Length 511;
Best Local Similarity 31.6%: Pred. No. 2.30e-01;
Matches 30; Conservative 26; Mismatches 32; Indels 7; Gaps 6;

Db 321 IAAVHNYAA-IASGMVRAGADITAIIDLRGTGTAAPRMIRDNYGIPVELALAAVDRLRD 379
   ||::| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 172 IMAGNVYTGKVEEELISGADITKV-GIGPGSVCTTRK-KTGVGYR-QLS-AVMECADAA 227
   ||::| | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 380 EGIRNKASILVAGGFRCSADVVRKAALGADAVYIG 414
   ||::| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 228 HGL--KGHIISDGGCSCPGDVAKAFAGADFYMLG 260
   ||::| | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: Mon Jun 15 16:02:12 1998
Job time : 82 secs.